Probab. Theory Relat. Fields (2013) 155:127–164
DOI 10.1007/s00440-011-0394-z

A solvable mixed charge ensemble on the line: global results

Brian Rider · Christopher D. Sinclair · Yuan Xu

Received: 19 July 2010 / Revised: 10 October 2011 / Published online: 27 October 2011

Abstract We consider an ensemble of interacting charged particles on the line consisting of two species of particles with charge ratio 2:1 in the presence of an external field. With the total charge fixed and the system held at temperature corresponding to $\beta = 1$, it is proved that the particles form a Pfaffian point process. When the external field is quadratic (the harmonic oscillator potential), we produce the explicit family of skew-orthogonal polynomials necessary to simplify the related matrix kernels. In this setting a variety of limit theorems are proved on the distribution of the number as well as the spatial density of each species of particle as the total charge increases to infinity. Connections to Ginibre's real ensemble of random matrix theory are highlighted throughout.

Keywords Random matrix · Eigenvalue statistics · Pfaffian processes

Mathematics Subject Classification (2000) 60B20 · 82B05

1 Introduction

Recent progress in the solvability of Ginibre's real ensemble (square matrices with real i.i.d. standard normal entries [9]) [4,8,15] sheds light on our understanding of ensembles with two different species of eigenvalues. Here we introduce a charged

B. Rider (✉)
Department of Mathematics, University of Colorado, Boulder, CO 80309, USA
e-mail: brian.rider@colorado.edu

C. D. Sinclair · Y. Xu
Department of Mathematics, University of Oregon, Eugene, OR 97403, USA
e-mail: csinclai@uoregon.edu

Y. Xu
e-mail: yuan@uoregon.edu

particle model, based loosely on Ginibre's real ensemble, but with particles of charge 2 on the line replacing the complex conjugate pairs of eigenvalues. The solvability of the ensemble (i.e. the existence of a Pfaffian point process on the particles) follows from a similar analysis to that in [4], though new global phenomena arise which do not appear in Ginibre's real ensemble.

Let L, M and N be non-negative integers so that $L + 2M = N$, and consider 1-dimensional (log-potential) electrostatic system consisting of L particles with unit charge and M particles with charge 2. We will identify the state of the system by pairs of finite subsets of $\mathbb{R}$, $\xi_1 = \{\alpha_1, \alpha_2, \ldots, \alpha_L\}$ and $\xi_2 = \{\beta_1, \beta_2, \ldots \beta_M\}$, where $\alpha_1, \alpha_2, \ldots, \alpha_L$ represent the locations of the charge 1 particles and $\beta_1, \beta_2, \ldots, \beta_M$ represent the locations of the charge 2 particles.

The potential energy of state $\xi = (\xi_1, \xi_2)$ is given by

$$\sum_{j<k} \log|\alpha_j - \alpha_k| + 4\sum_{m<n} \log|\beta_m - \beta_n| + 2\sum_{\ell=1}^{L}\sum_{m=1}^{M} |\alpha_\ell - \beta_m|.$$

We assume that the system is in the presence of an external field, so that the interaction energy between the charges and the field is given by

$$-\sum_{\ell=1}^{L} V(\alpha_\ell) - 2\sum_{m=1}^{M} V(\beta_m)$$

for some potential $V : \mathbb{R} \to [0, \infty)$. Eventually we will specify to the situation where V is the harmonic oscillator potential, but for now we maintain generality. The total potential energy of the system is therefore

$$\begin{aligned} E = &\sum_{j<k} \log|\alpha_j - \alpha_k| + 4\sum_{m<n} \log|\beta_m - \beta_n| + 2\sum_{\ell=1}^{L}\sum_{m=1}^{M} \log|\alpha_\ell - \beta_m| \\ &- \sum_{\ell=1}^{L} V(\alpha_\ell) - 2\sum_{m=1}^{M} V(\beta_m). \end{aligned} \tag{1.1}$$

Given a pair of vectors $(\boldsymbol{\alpha}, \boldsymbol{\beta}) \in \mathbb{R}^L \times \mathbb{R}^M$ we will define $E(\boldsymbol{\alpha}, \boldsymbol{\beta})$ to be the right hand side of (1.1), and call $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ a state *vector* corresponding to the state ξ. The number of relabellings of the particles is generically $L!M!$.

Assuming the system is placed in a heat bath corresponding to inverse temperature parameter $\beta = 1$, then the Boltzmann factor for the state vector $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ is given by

$$e^{-E(\boldsymbol{\alpha},\boldsymbol{\beta})} = \prod_{\ell=1}^{L} w(\alpha_\ell) \prod_{m=1}^{M} w(\beta_m)^2 \prod_{j<k} |\alpha_j - \alpha_k| \prod_{m<n} |\beta_m - \beta_n|^4 \prod_{\ell=1}^{L}\prod_{m=1}^{M} |\alpha_\ell - \beta_m|^2, \tag{1.2}$$

where $w(\gamma) = e^{-V(\gamma)}$ is the one-body Boltzmann factor due to the external field. The partition function of the system is given by

$$Z_{L,M} = \frac{1}{L!M!} \int_{\mathbb{R}^L} \int_{\mathbb{R}^M} e^{-E(\boldsymbol{\alpha},\boldsymbol{\beta})} d\mu^L(\alpha)\, d\mu^M(\beta), \tag{1.3}$$

where μ and μ^L are Lebesgue measure on $\mathbb{R}$ and $\mathbb{R}^L$ respectively. The multiplicative prefactor $1/(L!M!)$ compensates for the multitude of state vectors associated to each state.

Here we will be interested in a form of the grand canonical ensemble conditioned so that the sum of the charges equals N. That is, we consider the union of all two component ensembles with L particles of charge 1 and M particles of charge 2 over all pairs of non-negative integers L and M with $L + 2M = N$. The partition function of this ensemble is given by

$$Z(X) = \sum_{(L,M)} X^L Z_{L,M} = \sum_{(L,M)} \frac{X^L}{L!M!} \int_{\mathbb{R}^L} \int_{\mathbb{R}^M} e^{-E(\boldsymbol{\alpha},\boldsymbol{\beta})} d\mu^L(\alpha)\, d\mu^M(\beta).$$

Here $X \geq 0$ is the *fugacity* of the system, a parameter which controls the probability that the system has a particular population vector (L, M). The sum over (L, M) indicates that we are summing over all pairs of non-negative integers such that $L + 2M = N$.

Note now that (L, M) is itself a random vector, though we will continue to use this notation for the value of the population vector as well. For example, for each admissible pair (L, M), the joint density of particles given population vector (L, M) is given by the normalized Boltzmann factor,

$$\frac{X^L}{Z(X)} e^{-E(\boldsymbol{\alpha},\boldsymbol{\beta})}. \tag{1.4}$$

When $X = 1$ the probability of seeing a particular pair (L, M), or $\mathrm{Prob}(L, M)$, is the ratio $Z_{L,M}/Z$, where $Z = Z(1)$. We will focus much of our attention on the case $X = 1$ due to the analogy with the real Ginibre ensemble: the charge 1 particles play the role of real eigenvalues in real Ginibre and the charge 2 particles the role of pairs of complex conjugate eigenvalues which have been forced together on the real axis. As already mentioned, the solvability of the ensemble (that is the presence of a Pfaffian point process on the particles) follows from practically the same argument as in the real Ginibre ensemble. This observation was our original motivation for the present work.

Experts of random matrix theory will have already noticed that when $X = 0$ the above reduces to a general orthogonal (or $\beta = 1$) ensemble. Likewise, as $X \to \infty$, the above formally goes over to the corresponding symplectic (or $\beta = 4$) ensemble. Thus, the two-charge ensemble provides an interpolation between two classical and well studied point processes. Forrester previously introduced and studied the two-charge ensemble constrained to the circle with uniform weight, obtaining some similar results to the present (see §6.7 of [7] and the references therein).

2 Statistics of the particles

Here we are primarily concerned with global statistics of the particles when the potential V is quadratic,

$$V(\gamma) = \gamma^2/2, \qquad \text{or,} \qquad w(\gamma) = e^{-\gamma^2/2},$$

and the fugacity $X = 1$. Many of the results presented here are valid for other potentials and other values of X, as will be indicated in the appropriate places below. Throughout we restrict ourselves to the situation where $N = 2J$ is an even integer.

The rest of this section describes global behavior of the distribution of L (and M) as well as the global spatial distribution of each species of particles. These results hinge on being able to derive a Pfaffian point process for the particles, as well as the connected skew-orthogonal polynomials which allow a simplified formula for the matrix kernel for the process. These facts are presented in Sect. 3; proofs are contained in Sect. 4. The local analysis of this kernel (i.e. its scaling limits in the bulk and at the edge) as well as investigation of the right-most particle of each species will appear in a forthcoming publication.

2.1 Distribution of population vectors

Sharp results on the law of the state vector (L, M) are consequences of the following characterization in terms of generalized Laguerre polynomials. Recall that the generalized Laguerre polynomials with parameter α are orthogonal with respect to $x^\alpha e^{-x}$ on $[0, \infty)$. For our purposes, we set $\alpha = -1/2$ and define L_j to be the degree j polynomial which satisfies

$$\int_0^\infty x^{-1/2} e^{-x} L_j(x) L_k(x)\, dx = \delta_{j,k} \frac{\Gamma\left(j + \frac{3}{2}\right)}{j!}.$$

Theorem 2.1 *When the fugacity is set equal to one,* $\text{Prob}(L, M)$ *is the coefficient of* X^L *of the polynomial* $L_{N/2}(-X^2)/L_{N/2}(-1)$. *That is,*

$$\frac{Z(X)}{Z} = \frac{L_{N/2}(-X^2)}{L_{N/2}(-1)},$$

and so

1. $\text{Prob}(L, M) = \dfrac{2^L}{L!M!} \Big/ \left(\sum_{(\ell,m)\ell+2m=N} \dfrac{2^\ell}{\ell! m!} \right)$ *if L is even, and is equal to 0 otherwise,*
2. $\mathbb{E}(L^m) = \left[(X\dfrac{d}{dX})^m \dfrac{L_{N/2}(-X^2)}{L_{N/2}(-1)} \right]_{X=1}$ *for m non-negative integer.*

Properties of the Laguerre polynomials now allow for nice expressions for the mean, variance, etc. of L for all finite values of N. For example, we have that

$$\mathbb{E}[L] = \frac{d}{dX}\left[\frac{Z(X)}{Z}\right]_{X=1} = 2\sum_{j=0}^{N/2-1}\frac{L_j(-1)}{L_{N/2}(-1)}$$
$$= \frac{2\sum_{i=0}^{N/2-1}\left[\Gamma\left(\frac{N}{2}-i\right)\Gamma\left(i+\frac{3}{2}\right)i!\right]^{-1}}{\sum_{i=0}^{N/2}\left[\Gamma\left(\frac{N}{2}-i+1\right)\Gamma\left(i+\frac{1}{2}\right)i!\right]^{-1}}.$$

Asymptotic descriptions of the law of L are just as readily obtained from Theorem 2.1.

Theorem 2.2 *For $X = 1$ and $N \to \infty$ it holds*:

1. $\mathbb{E}(L) = \sqrt{2N} - 1 + \frac{1}{3\sqrt{N}} + \mathrm{O}(N^{-1})$ *and* $\mathrm{Var}(L) = \sqrt{2N} - \frac{4}{3} + \mathrm{O}(N^{-1/2})$,
2. $\dfrac{L-(2N)^{1/2}}{(2N)^{1/4}}$ *converges in distribution to a standard Normal random variable,*
3. $\mathrm{Prob}\left(|\frac{L}{\sqrt{2N}} - 1| \geq \epsilon\right) \leq CNe^{-(\epsilon\wedge 1)\sqrt{2N}}$ *with a numerical constant C for any $\epsilon > 0$.*

Spelling out a few higher order terms for the mean and variance gives an example of what is possible given the explicit Laguerre polynomial formulas. Going out further in the expansions is possible; naturally one obtains one less order in the variance than the mean. More importantly, the number of real eigenvalues in the real Ginibre ensemble also has both mean and variance of $O(\sqrt{N})$ (see [6,8]), providing another point of contact between these ensembles. On the other hand, looking through the proof of Theorem 2.2 shows that the left tail of the distribution of L has the so-called mean-field form. In particular, the leading order of this tail may be expressed as $\exp(-\sqrt{N}f(t/\sqrt{N}))$ for $f(t) = 1 - t\log t - t$. This may be taken as an indication that the fluctuating species (here the charge 1 points) become uncorrelated, see the discussion in [11]. In fact, in the corresponding circular ensemble due to Forrester (again, [7] §6.7), it has been shown that the (bulk) connected 2-point function for the charge 1 points vanishes in the limit. The same does not hold for the real eigenvalues in real Ginibre, so despite the obvious analogies with our $X = 1$ ensemble there remain important differences.

It is also worth pointing out that the shape of the tail (the form of $f(t)$) of the number of real eigenvalues in the real Ginibre ensemble is not known. Whether or not that number possesses a central limit theorem is also an open question.

2.2 Spatial density of particles

Introduce the (mean) counting measures ρ_1 and ρ_2 for the charge 1 and charge 2 particles defined by

$$\mathbb{E}\left[|A\cap\xi_1|\right] = \int_A d\rho_1 \quad \text{and} \quad \mathbb{E}\left[|A\cap\xi_2|\right] = \int_A d\rho_2$$

for Borel subsets $A \subseteq \mathbb{R}$ (where, for instance, $|A \cap \xi_1|$ is the number of charge 1 particles in A). As we shall see in the sequel, these measures are absolutely continuous with respect to Lebesgue measure; we write $R^{(N)}_{1,0}(x)$ and $R^{(N)}_{0,1}(x)$ for their respective densities. (The cryptic notation will be resolved in Sect. 3.2, when we define the ℓ, m-correlation function of the ensemble to be $R^{(N)}_{\ell,m}$).

Again keeping $X = 1$, from Theorem 2.2 we see that, as $N \to \infty$,

$$\int_{-\infty}^{\infty} R^{(N)}_{1,0}(x)\,dx = \mathbb{E}[L] \sim \sqrt{2N}, \quad \text{and} \quad \int_{-\infty}^{\infty} R^{(N)}_{0,1}(x)\,dx = \mathbb{E}[M] \sim \frac{N}{2}.$$

One then would ask, when suitably scaled and normalized as in

$$s^{(N)}_1(x) = \frac{1}{\sqrt{2}} R^{(N)}_{1,0}(\sqrt{N}x) \quad \text{and} \quad s^{(N)}_2(x) = \frac{2}{\sqrt{N}} R^{(N)}_{0,1}(\sqrt{N}x), \tag{2.1}$$

whether $s^{(N)}_1(x)dx$ and $s^{(N)}_2(x)dx$ converge to proper probability measures.

Theorem 2.2 shows that, with probability one, for all N large the number of charge 1 particles is $\sqrt{2N}(1+o(1))$, suggesting that the limiting statistics of the charge 2 particles should behave as though there are no charge 1 particles present. (Or like a copy of the Gaussian Symplectic Ensemble, again arrived at from the present ensemble upon setting $L = 0$). Indeed we find the scaled density of charge 2 particles approaches the semicircle law.

Contrariwise, though the charge 1 particles exhibit (at finite N) the same level repulsion amongst themselves as the eigenvalues in the Gaussian Orthogonal Ensemble (occurring here when $M = 0$), the preponderance of charge 2 particles leads to a different limit distribution. This turns out to be the same (up to constant scalings) as that for the real eigenvalues in Ginibre's real ensemble ([4]), i.e., the uniform distribution.

Theorem 2.3 *As defined in (2.1), $s^{(N)}_1$ converges weakly in the sense of measures to the uniform law on $[-\sqrt{2}, \sqrt{2}]$, and $s^{(N)}_2$ converges in the same manner to the semi-circular law with the same support. In particular,*

$$\lim_{N\to\infty} \int e^{itx} s^{(N)}_1(x)dx = \frac{1}{\sqrt{2}t} \sin(\sqrt{2}t),$$

and

$$\lim_{N\to\infty} \int e^{itx} s^{(N)}_2(x)dx = \frac{\sqrt{2}}{t} J_1(\sqrt{2}t).$$

Here J_1 is the Bessel function of the first kind, and the point is that $\int_{-\sqrt{2}}^{\sqrt{2}} e^{itx}\sqrt{2-x^2}\,\frac{dx}{\pi} = \frac{\sqrt{2}}{t} J_1(\sqrt{2}t)$. We give an elementary proof of Theorem 2.3, making use of the explicit skew-orthogonal polynomial system derived below. Given that the number of charge 1 particles is almost surely $o(N)$, one could undoubtedly make a large deviation

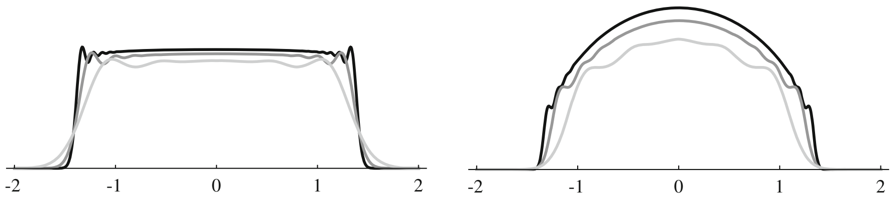

Fig. 1 $s_1^{(N)}$ (*left*) and $s_2^{(N)}$ (*right*) for, from lightest to darkest, $N = 10$, 30 and 90

proof along the lines of [2] or [3] of a stronger version of the second statement: that the random counting measure of charge 2 particles converges almost surely to the semi-circle law. However, it is not immediately clear how to use such energy optimization ideas to access the charge 1 profile (Fig. 1).

2.3 Tuning the fugacity

Again, the focus on the case $X = 1$ stems from the transparent connections to the real Ginibre ensemble. It is also clear that the asymptotic appraisals of Theorems 2.2 and 2.3 will remain unchanged for any fixed X as $N \to \infty$, save the adjustment of constants (Fig. 2).

Other ranges of X remain interesting. In particular it is natural to ask for what $X = X(N)$ are the number of charge 1 and charge 2 particles of the same order.

Theorem 2.4 *Let* $X = \sqrt{N\gamma}$ *for* $\gamma > 0$. *Then,*

$$\mathbb{E}[L] = N\gamma(\sqrt{1+2/\gamma} - 1)(1+o(1)), \qquad \mathrm{Var}[L] = N\gamma(\sqrt{1+2/\gamma}-1)(1+o(1))$$

as $N \to \infty$.

We have not considered further statistics of L (such as an actual limit theorem) in this regime, nor do we spell out the higher order corrections to the mean and variance (though they are relatively easy to obtain). Do note though that $\gamma \mapsto \gamma(\sqrt{1+2/\gamma}-1)$ is strictly increasing from zero (as $\gamma \downarrow 0$) to one (as $\gamma \uparrow \infty$), so the mean takes values over its full possible range.

That $X = O(\sqrt{N})$ is enough to move the average number of particles (of charge 1) while X only enters the ensemble average in a polynomial fashion is rather intriguing, particularly again from a large deviations viewpoint through which one might hope to prove almost sure convergence of the spatial densities to a maximizer of $Z(X)$. Just as in the $X = 1$ setting we content ourselves here with convergence of the mean densities, via characteristic functions. It is of course still possible that the mean spatial densities still converge. Similar to (2.1) set

$$s_1^{(N,\gamma)}(x) = \frac{1}{\sqrt{N}} R_{1,0}^{(N)}(\sqrt{N}x) \quad \text{and} \quad s_2^{(N,\gamma)}(x) = \frac{1}{\sqrt{N}} R_{0,1}^{(N)}(\sqrt{N}x).$$

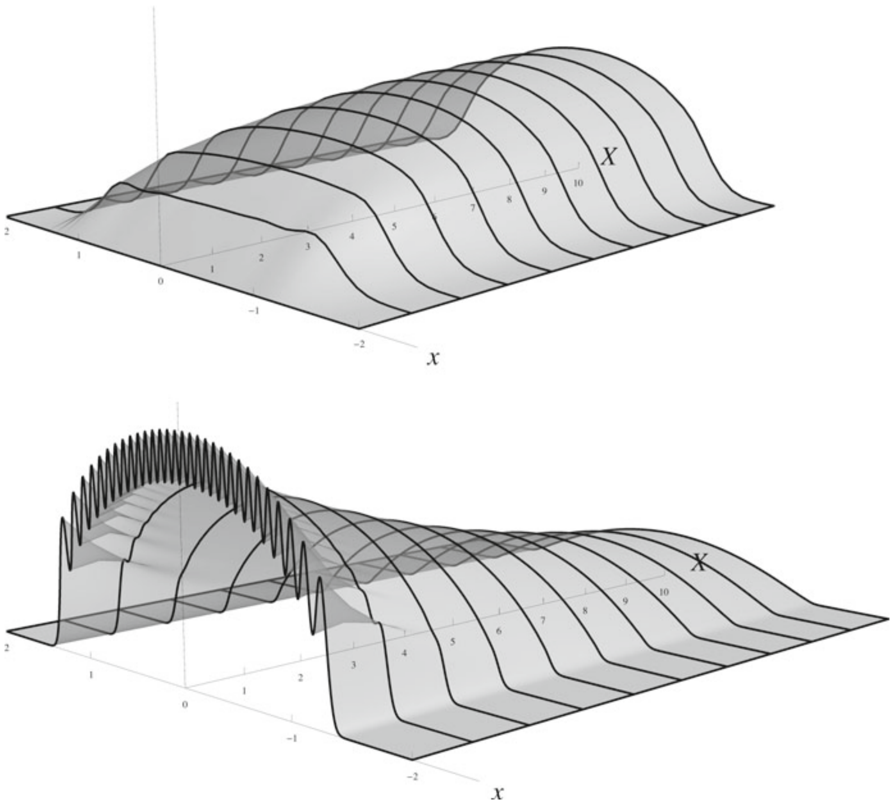

Fig. 2 $s_1^{(60)}$ (*above*) and $s_2^{(60)}$ (*below*) as a function of the fugacity X for the range $0 \le X \le 10$

The added superscript γ serves as a reminder that $X = \sqrt{N\gamma}$ is implicit in the definitions on the right hand side. To somewhat ease the notation we have not included normalizers which would make any limiting $s_{1,2}^{(N,\gamma)}$ have mass one.

Theorem 2.5 *Define* $u_\gamma(x) = \sqrt{1 + 2x/\gamma}$ *and* $v_\gamma(x) = \frac{x}{x+\gamma(1+u_\gamma(x))}$. *Then,*

$$\lim_{N\to\infty} \int e^{itx} s_1^{(N,\gamma)}(x)dx = \int_0^1 \sum_{m=0}^{\infty} \frac{(-t^2x/2)^m}{(m!)^2} \frac{dx}{u_\gamma(x)} + \int_0^1 \sum_{m=1}^{\infty} (-t^2x/2)^m \sum_{\ell=1}^{m} \frac{(v_\gamma(x))^\ell}{(m-\ell)!(m+\ell)!} \frac{dx}{u_\gamma(x)}. \tag{2.2}$$

Denoting the right hand side by $\sigma_\gamma(t)$, $\lim_{N\to\infty} \int e^{itx} s_2^{(N,\gamma)}(x)dx = \frac{1}{\sqrt{2}t} J_1(\sqrt{2}t) - \frac{1}{2}\sigma_\gamma(t)$.

We have not succeeded in inverting $\sigma_\gamma(t)$, but here are a few simple observations. After normalizing by $\gamma^{-1}(\sqrt{1+2/\gamma}-1)^{-1}$, the first part (line one of the right hand side of (2.2) has the interpretation of (the characteristic function of) an arcsine random variable times the square root of an independent random variable with density proportional to $u_\gamma(x)^{-1}$. To wit, this term may be rewritten as in $\int_0^1 J_0(\sqrt{2x}\, t)\frac{dx}{u_\gamma(x)}$. Since $u_\gamma(x) \to 1$ and $v_\gamma(x) \to 0$ as $\gamma \uparrow \infty$, this gives back a well-known characterization of the semicircle law. Likewise, as $\gamma \downarrow 0$ simplifications arise from $v_\gamma(x) \to 1$, and we have that $\gamma^{-1}(\sqrt{1+2/\gamma}-1)^{-1}\sigma_\gamma(t)$ tends to $\frac{1}{\sqrt{2}t}\sin(\sqrt{2}t)$. In other words, the not as of yet explicit limit density does interpolate between the uniform and the semicircle laws. Lastly, the content of the second statement of Theorem 2.5 is that the limit of the "total" density $s_1^{(N,\gamma)} + 2s_2^{(N,\gamma)}$ is semicircle.

3 A Pfaffian point process for the particles

All of the results in this paper follow, in one way or another, from the fact that our interacting particles form a Pfaffian point process very much like that of Ginibre's real ensemble and related to the Gaussian Orthogonal and Symplectic Ensembles.

The results in this section are valid for quite general weight functions w and fugacities. Thus, for the time being, we will return to the general situation.

3.1 The joint density of particles

The joint density of particles for a particular choice of (L, M) is given by

$$\frac{1}{Z(X)}\frac{X^L}{L!M!}\Omega_{L,M}(\boldsymbol{\alpha}, \boldsymbol{\beta}), \qquad \text{where} \qquad \Omega_{L,M}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = e^{-E(\boldsymbol{\alpha},\boldsymbol{\beta})}.$$

More specifically,

$$\Omega_{L,M}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \prod_{\ell=1}^{L} w(\alpha_\ell) \prod_{m=1}^{M} w(\beta_m)^2 \prod_{j<k} |\alpha_j - \alpha_k| \prod_{m<n} |\beta_m - \beta_n|^4 \prod_{\ell=1}^{L}\prod_{m=1}^{M} |\alpha_\ell - \beta_m|^2;$$

where, for now, the only assumptions we will make on w are that it is positive and Lebesgue measurable with $0 < Z(X) < \infty$.

3.2 Correlation functions

Given $0 \le \ell \le L$ and $0 \le m \le M$, we define the ℓ, m-correlation function $R_{\ell,m}^{(N)} : \mathbb{R}^\ell \times \mathbb{R}^m \to [0, \infty)$ by

$$R_{\ell,m}^{(N)}(\mathbf{x}; \mathbf{y}) = \frac{1}{Z(X)} \sum_{\substack{(L,M)\\ L\ge\ell, M\ge m}} \frac{X^L}{(L-\ell)!(M-m)!} \times \int_{\mathbb{R}^{L-\ell}}\int_{\mathbb{R}^{M-m}} \Omega_{L,M}(\mathbf{x}\vee\boldsymbol{\alpha}, \mathbf{y}\vee\boldsymbol{\beta})\, d\mu^{L-\ell}(\boldsymbol{\alpha})\, d\mu^{M-m}(\boldsymbol{\beta}),$$

where, for instance, $\mathbf{x} \vee \boldsymbol{\alpha}$ is the vector in $\mathbb{R}^L$ formed by concatenating $\mathbf{x} \in \mathbb{R}^\ell$ and $\boldsymbol{\alpha} \in \mathbb{R}^{L-\ell}$. We will often write $R_{\ell,m}$ for $R^{(N)}_{\ell,m}$ in situations where N is seen as being fixed.

The correlation functions encode statistical information about the configurations of the charged particles. To be more precise, given $\boldsymbol{\alpha} \in \mathbb{R}^L$ and $\boldsymbol{\beta} \in \mathbb{R}^M$ with $L + 2M = N$, we set

$$\xi = \xi(\boldsymbol{\alpha}, \boldsymbol{\beta}) = (\xi_1, \xi_2) = \big(\xi_1(\boldsymbol{\alpha}), \xi_2(\boldsymbol{\beta})\big) = \big(\{\alpha_1, \dots, \alpha_\ell\}, \{\beta_1, \dots \beta_m\}\big).$$

Given an L-tuple of mutually disjoint subsets of $\mathbb{R}$, $\mathbf{A} = (A_1, A_2, \dots, A_L)$, and an M-tuple of mutually disjoint subsets of $\mathbb{R}$, $\mathbf{B} = (B_1, B_2, \dots, B_M)$, the probability that the system is in a state where there is exactly one charge 1 particle in each of the A_ℓ and exactly one charge 2 particle in each of the B_m is given by

$$\mathbb{E}\left[\left\{\prod_{\ell=1}^{L} |A_\ell \cap \xi_1|\right\}\left\{\prod_{m=1}^{M} |B_m \cap \xi_2|\right\}\right] = \frac{X^L}{Z(X)} \int_{\mathbf{B}}\int_{\mathbf{A}} \Omega_{L,M}(\boldsymbol{\alpha}, \boldsymbol{\beta})\, d\mu^L(\boldsymbol{\alpha})\, d\mu^M(\boldsymbol{\beta}).$$

The correlation functions can be used to generalize this formula. If $\mathbf{A} = (A_1, A_2, \dots, A_\ell)$ is a tuple of disjoint subsets of $\mathbb{R}$ and $\mathbf{B} = (B_1, B_2, \dots, B_m)$ another such tuple, then

$$\mathbb{E}\left[\left\{\prod_{j=1}^{\ell} |A_j \cap \xi_1|\right\}\left\{\prod_{k=1}^{m} |B_k \cap \xi_2|\right\}\right] = \int_{\mathbf{B}}\int_{\mathbf{A}} R_{\ell,m}(\mathbf{x}; \mathbf{y})\, d\mu^\ell(\boldsymbol{\alpha})\, d\mu^m(\boldsymbol{\beta}). \tag{3.1}$$

Here we are assuming ℓ and m are non-negative integers with $\ell + 2m \le N$. The origin of the sum in the definition of $R_{\ell,m}$ is now clear; our choice of $\mathbf{A}$ and $\mathbf{B}$ no longer specify a single value of the population vector (L, M) and we have to sum over all population vectors which contribute to the expectation on the left hand side of (3.1).

3.3 Pfaffian point processes

Consider, for the moment, a simplified system of indistinguishable random points $\zeta = \{\gamma_1, \gamma_2, \dots, \gamma_N\} \subseteq \mathbb{R}$ with correlation functions $R_n(\mathbf{z})$ satisfying

$$\mathbb{E}\left[\prod_{j=1}^{n} |A_j \cap \zeta|\right] = \int_{A_1} \cdots \int_{A_n} R_n(\mathbf{z})\, d\mu^n(\mathbf{z})$$

for any n-tuple $(A_1, A_2, \dots, A_n)$ of mutually disjoint sets.

The Pfaffian of an antisymmetric $2n \times 2n$ matrix $\mathbf{K} = [k_{j,k}]$, is defined by

$$\operatorname{Pf} \mathbf{K} = \frac{1}{n!2^n} \sum_{\sigma \in S_{2n}} \operatorname{sgn} \sigma \prod_{j=1}^{n} k_{\sigma(2n-1),\sigma(2n)},$$

where, as is usual, S_{2n} is the symmetric group on $2n$ numbers. Though not obvious from this formula, the Pfaffian and determinant are related by the important identity $\det \mathbf{K} = (\operatorname{Pf} \mathbf{K})^2$.

If there exists a matrix valued function $K_N : \mathbb{R}^2 \to \mathbb{R}^{2\times 2}$ such that

$$R_n(\mathbf{z}) = \operatorname{Pf}\left[K_N(z_j, z_k)\right]_{j,k=1}^{n},$$

then we say that our ensemble of random points forms a *Pfaffian point process* with *matrix kernel* K_N. Much of the information about probabilities of locations of particles (e.g. gap probabilities) can be derived from properties of the matrix kernel. Moreover, in many instances, we are interested in statistical properties of the particles as their number (or some related parameter) tends toward ∞. In these instances, it is sometimes possible to analyze $K_N(x, y)$ in this limit (under, perhaps, some scaling of x and y dependent on N) so that the relevant limiting probabilities are attainable from this limiting kernel.

For the ensemble of charge 1 and charge 2 particles with total charge N, we will demonstrate that the correlation functions have a Pfaffian formulation of the form,

$$R_{\ell,m}(\mathbf{x}; \mathbf{y}) = 2^{\ell} \operatorname{Pf}\begin{bmatrix} K_N^{1,1}(x_j, x_{j'}) & K_N^{1,2}(x_j, z_{k'}) \\ K_N^{2,1}(z_k, x_{k'}) & K_N^{2,2}(x_k, x_{k'}) \end{bmatrix}; \qquad \begin{matrix} j, j' = 1, 2, \ldots, \ell \\ k, k' = 1, 2, \ldots m \end{matrix} \tag{3.2}$$

where $K_N^{1,1}$, $K_N^{1,2}$, $K_N^{2,1}$ and $K_N^{2,2}$ are 2×2 matrix kernels.

3.4 A Pfaffian form for the total partition function

In order to establish the existence of the matrix kernels we first need a Pfaffian formulation of the total partition function.

Given a measure ν on $\mathbb{R}$ we define the operators ϵ_1^{ν} and ϵ_4^{ν} on $L^2(\nu)$ by

$$\epsilon_1^{\nu} f(x) = \frac{1}{2}\int_{\mathbb{R}} f(y)\operatorname{sgn}(y - x)d\nu(y) \qquad \text{and} \qquad \epsilon_4^{\nu} f(y) = f'(y).$$

(Obviously ϵ_4^{ν} does not depend on ν, but it is convenient to maintain symmetric notation). Using these inner products we define

$$\langle f|g\rangle_{\beta}^{\nu} = \int_{\mathbb{R}} \left[f(x)\epsilon_{\beta} g(x) - g(x)\epsilon_{\beta} f(x)\right] d\nu(x), \qquad \beta = 1, 4.$$

We specialize these operators and inner products for Lebesgue measure μ by setting $\epsilon_{\beta} = \epsilon_{\beta}^{\mu}$ and $\langle f|g\rangle_{\beta}^{\mu}$. We also write $\widetilde{f}(x) = w(x)f(x)$. It is easily seen that

$$\langle \widetilde{f}|\widetilde{g}\rangle_1 = \int_{\mathbb{R}} \left[\widetilde{f}(x)\epsilon_1\widetilde{g}(x) - \widetilde{g}(x)\epsilon_1\widetilde{f}(x)\right] d\mu(x) = \langle f|g\rangle_1^{w\mu}.$$

Similarly,

$$\langle \widetilde{f} | \widetilde{g} \rangle_4 = \int_{\mathbb{R}} \left[\widetilde{f}(x) \frac{d}{dx} \widetilde{g}(x) - \widetilde{g}(x) \frac{d}{dx} \widetilde{f}(x) \right] d\mu(x)$$
$$= \int_{\mathbb{R}} w(x)^2 \left[f(x) g'(x) - g(x) f'(x) \right] dx = \langle f | g \rangle_4^{w^2 \mu}.$$

We call a family of polynomials, $\mathbf{p} = \big(p_0(x), p_1(x), \ldots, p_{N-1}(x)\big)$, a *complete* family of polynomials if $\deg p_n = n$. A complete family of monic polynomials is defined accordingly.

Theorem 3.1 *Suppose N is even and* $\mathbf{p}$ *is any complete family of monic polynomials. Then,*

$$Z(X) = \operatorname{Pf}\left(X^2 \mathbf{A}^{\mathbf{p}} + \mathbf{B}^{\mathbf{p}} \right),$$

where

$$\mathbf{A}^{\mathbf{p}} = [\langle \widetilde{p}_m | \widetilde{p}_n \rangle_1]_{m,n=0}^{N-1} \quad \textit{and} \quad \mathbf{B}^{\mathbf{p}} = [\langle \widetilde{p}_m | \widetilde{p}_n \rangle_4]_{m,n=0}^{N-1}.$$

Corollary 3.2 *With the same assumptions as Theorem 3.1, $Z = \operatorname{Pf}(\mathbf{A}^{\mathbf{p}} + \mathbf{B}^{\mathbf{p}})$.*

3.5 A Pfaffian formulation of the correlation functions

In order to describe the entries in the kernels $K_N^{1,1}$, $K_N^{1,2}$, $K_N^{2,1}$ and $K_N^{2,2}$ appearing in (3.2), we suppose $\mathbf{p}$ is any complete family of polynomials and define

$$\mathbf{C}^{\mathbf{p}} = X^2 \mathbf{A}^{\mathbf{p}} + \mathbf{B}^{\mathbf{p}},$$

where $\mathbf{A}^{\mathbf{p}}$ and $\mathbf{B}^{\mathbf{p}}$ are as in Corollary 3.2. Since we are assuming that $Z = \operatorname{Pf} \mathbf{C}^{\mathbf{p}}$ is non-zero, $\mathbf{C}^{\mathbf{p}}$ is invertible and we set the inverse transpose of $\mathbf{C}^{\mathbf{p}}$ to be

$$(\mathbf{C}^{\mathbf{p}})^{-\mathsf{T}} = \left[\zeta_{j,k} \right]_{j,k=0}^{N-1}.$$

The $\zeta_{j,k}$ clearly depend on our choice of polynomials. We then define

$$\varkappa_N(x, y) = \sum_{j,k=0}^{N-1} \widetilde{p}_j(x) \zeta_{j,k} \widetilde{p}_k(y). \tag{3.3}$$

The operators ϵ_1 and ϵ_4 operate on $\varkappa_N(x, y)$ in the usual manner. For instance,

$$\epsilon_4 \varkappa_N(x, y) = \sum_{j,k=0}^{N-1} \epsilon_4 \widetilde{p}_j(x) \zeta_{j,k} \widetilde{p}_k(y)$$

and

$$\varkappa_N \epsilon_1(x, y) = \sum_{j,k=0}^{N-1} \widetilde{p}_j(x)\zeta_{j,k}\epsilon_1\widetilde{p}_k(y).$$

(That is, ϵ written on the left acts on the $\varkappa_N(x, y)$ viewed as a function of x, etc.).

Theorem 3.3 *Suppose N is even,* $\mathbf{p}$ *is any complete family of polynomials and $\varkappa_N(x, y)$ is given as in* (3.3)*. Then,*

$$R_{\ell,m}(\mathbf{x}; \mathbf{y}) = 2^\ell \operatorname{Pf}\begin{bmatrix} K_N^{1,1}(x_j, x_{j'}) & K_N^{1,2}(x_j, y_{k'}) \\ K_N^{2,1}(y_k, x_{k'}) & K_N^{2,2}(y_k, y_{k'}) \end{bmatrix}; \qquad \begin{matrix} j, j' = 1, 2, \dots, \ell \\ k, k' = 1, 2, \dots m, \end{matrix} \tag{3.4}$$

where

$$K_N^{1,1}(x, y) = \begin{bmatrix} \varkappa_N(x, y) & X^2\varkappa_N\epsilon_1(x, y) \\ X^2\epsilon_1\varkappa_N(x, y) & X^4\epsilon_1\varkappa_N\epsilon_1(x, y) + \frac{1}{4}\operatorname{sgn}(y - x) \end{bmatrix},$$
$$K_N^{2,2}(x, y) = \begin{bmatrix} \varkappa_N(x, y) & \varkappa_N\epsilon_4(x, y) \\ \epsilon_4\varkappa_N(x, y) & \epsilon_4\varkappa_N\epsilon_4(x, y) \end{bmatrix},$$
$$K_N^{1,2}(x, y) = \begin{bmatrix} \varkappa_N(x, y) & X^2\varkappa_N\epsilon_1(x, y) \\ \epsilon_4\varkappa_N(x, y) & X^2\epsilon_4\varkappa_N\epsilon_1(x, y) \end{bmatrix}$$

and

$$K_N^{2,1}(x, y) = \begin{bmatrix} \varkappa_N(x, y) & \varkappa_N\epsilon_4(x, y) \\ X^2\epsilon_1\varkappa_N(x, y) & X^2\epsilon_1\varkappa_N\epsilon_4(x, y) \end{bmatrix}.$$

When ℓ or m equal 0 the Pfaffian appearing on the right hand side of (3.4) *consists only of blocks formed from $K_N^{1,1}$ and $K_N^{2,2}$ respectively.*

Remark The factor 2^ℓ can be moved inside the Pfaffian so that the entries in the various kernels where an ϵ_1 appears are multiplied by 2. This maneuver is superficial, but has the effect of making these particular entries appear more like the entries in other $\beta = 1$ ensembles (e.g. GOE). For instance, $\frac{1}{2}\operatorname{sgn}(y - x)$ appears more natural to experts used to these other ensembles.

We notice in particular that the functions $R_{1,0}^{(N)}$ and $R_{0,1}^{(N)}$ given in Sect. 2.2 are given by

$$R_{1,0}^{(N)}(x) = 2X^2 \sum_{j,k=0}^{N-1} \widetilde{p}_j(x)\zeta_{j,k}\epsilon_1\widetilde{p}_k(x) \quad \text{and} \quad R_{0,1}^{(N)}(x) = \sum_{j,k=0}^{N-1} \widetilde{p}_j(x)\zeta_{j,k}\epsilon_4\widetilde{p}_k(x). \tag{3.5}$$

3.6 Skew-orthogonal polynomials

The entries in the kernel themselves can be simplified (or at least presented in a simplified form) by a judicious choice of **p**. If we define

$$\langle f|g\rangle^{(X)} = X^2\langle f|g\rangle_1 + \langle f|g\rangle_4,$$

then

$$\mathbf{C}^{\mathbf{p}} = \left[\langle \widetilde{p}_m|\widetilde{p}_n\rangle^{(X)}\right]_{m,n=0}^{N-1}.$$

Since $\varkappa_N$ (and by extension all other entries of the various kernels) depend on the inverse transpose of $\mathbf{C}^{\mathbf{p}}$, it is desirable to find a complete family of polynomials for which $\mathbf{C}^{\mathbf{p}}$ can be easily inverted.

We say $\mathbf{p} = (p_0, p_1, \ldots)$ is a family of *skew-orthogonal* polynomials for the skew-inner product $\langle\cdot|\cdot\rangle$ with weight w if there exists real numbers (called *normalizations*) $r_1, r_2, \ldots$ such that

$$\langle \widetilde{p}_{2j}|\widetilde{p}_{2k}\rangle = \langle \widetilde{p}_{2j+1}|\widetilde{p}_{2k+1}\rangle = 0 \qquad \text{and} \qquad \langle \widetilde{p}_{2j}|\widetilde{p}_{2k+1}\rangle = -\langle \widetilde{p}_{2k+1}|\widetilde{p}_{2j}\rangle = \delta_{j,k} r_j.$$

Using these polynomials, the entries in the matrix kernels presented in Sect. 3.5 have a particularly simple form. For instance,

$$\varkappa_N(x,y) = \sum_{j=0}^{J-1} \frac{\widetilde{p}_{2j}(x)\widetilde{p}_{2j+1}(y) - \widetilde{p}_{2j+1}(x)\widetilde{p}_{2j}(y)}{r_j},$$

and the entries of the kernels are computed by applying the appropriate ϵ operators to this expression.

3.7 Specification to the harmonic oscillator potential

We now return to the case where the weight function is $w(x) = e^{-x^2/2}$.

Theorem 3.4 *A complete family of skew-orthogonal polynomials for the weight w with respect to $\langle\cdot|\cdot\rangle^{(X)}$ is given by*

$$P_{2j}^{(X)}(x) = \sum_{k=0}^{j} (-1)^k \frac{L_k(-X^2)}{L_k(0)} L_k(x^2), \tag{3.6}$$

and

$$\begin{aligned} P_{2j+1}^{(X)}(x) &= 2xP_{2j}^{(X)}(x) - 2\frac{d}{dx}P_{2j}^{(X)}(x) \\ &= 4X^2 x \sum_{k=0}^{m-1} (-1)^k \frac{L_k^{(\frac{1}{2})}(-X^2)}{L_k^{(\frac{1}{2})}(0)} L_k^{(\frac{1}{2})}(x^2) + 2x(-1)^m \frac{L_m^{(-\frac{1}{2})}(-X^2)}{L_m^{(-\frac{1}{2})}(0)} L_m^{(\frac{1}{2})}(x^2). \end{aligned} \tag{3.7}$$

where $L_k(x) = L_k^{(-1/2)}(x)$ *is the generalized kth Laguerre polynomial. The normalization of this family of polynomials is given by*

$$\langle \widetilde{P}_{2m}^{(X)} | \widetilde{P}_{2m+1}^{(X)} \rangle^{(X)} = \frac{4\pi (m+1)!}{\Gamma\left(m+\frac{1}{2}\right)} L_m(-X^2) L_{m+1}(-X^2). \tag{3.8}$$

We can recover a family of monic skew-orthogonal polynomials by dividing by the leading coefficient. Specifically,

Corollary 3.5 *A complete family of monic skew-orthogonal polynomials for the weight* w *with respect to* $\langle \cdot | \cdot \rangle^{(X)}$ *is given by*

$$p_{2j}^{(X)}(x) = \frac{L_j(0)\, j!}{L_j(-X^2)} \sum_{k=0}^{j} (-1)^k \frac{L_k(-X^2)}{L_k(0)} L_k(x^2),$$

and

$$p_{2j+1}^{(X)}(x) = x p_{2j}^{(X)}(x) - \frac{d}{dx} p_{2j}^{(X)}(x).$$

The normalization for this family of monic skew-orthogonal polynomials is given by

$$r_j^{(X)} = \langle \widetilde{p}_{2j}^{(X)} | \widetilde{p}_{2j+1}^{(X)} \rangle^{(X)} = 4 \frac{(j+1)!\, \Gamma\left(j+\frac{1}{2}\right)}{j!} \frac{L_{j+1}(-X^2)}{L_j(-X^2)}.$$

Setting $X = 1$, we recover a family of skew-orthogonal polynomials for the harmonic oscillator two charge ensemble with fugacity equal to one, and we will write p_n for $p_n^{(1)}$ and r_j for $r_j^{(1)}$.

4 Proofs

4.1 Proof of Theorem 2.1

We set $J = N/2$. To prove 1, we use Theorem 3.1 and the skew-orthogonal polynomials from Corollary 3.5 to write

$$Z(X) = \mathrm{Pf} \begin{bmatrix} 0 & r_0^{(X)} & & & \\ -r_0^{(X)} & 0 & & & \\ & & \ddots & & \\ & & & 0 & r_{J-1}^{(X)} \\ & & & -r_{J-1}^{(X)} & 0 \end{bmatrix} = \prod_{j=0}^{J-1} r_j^{(X)}.$$

Hence,

$$\frac{Z(X)}{Z} = \prod_{j=0}^{J-1} \frac{r_j^{(X)}}{r_j} = \frac{L_J(-X^2)L_0(-1)}{L_J(-1)L_0(-X^2)} = \frac{L_J(-X^2)}{L_J(-1)},$$

where again $L_J(x) = L_J^{(-1/2)}(x)$. Note $L_0(X) = 1$.

The expression for $\mathrm{Prob}(L, M)$ then follows from explicit formulas for the corresponding Laguerre polynomials; the expression for $\mathbb{E}[L^m]$ is self-evident.

4.2 Proof of Theorem 2.2

Point 3 of Theorem 2.1 specified to the first two moments produces

$$\mathbb{E}[L] = \frac{d}{dX}\left[\frac{Z(X)}{Z}\right]_{X=1}, \quad \mathrm{Var}(L) = \left[\frac{d}{dX}\left(X\frac{d}{dX}\frac{Z(X)}{Z}\right) - \left(\frac{d}{dX}\frac{Z(X)}{Z}\right)^2\right]_{X=1}.$$

Now, since $L_J'(x) = -L_{J-1}^{1/2}(x)$ and $L_J(x) = L_J^{1/2}(x) - L_{J-1}^{1/2}(x)$, we have that

$$\mathbb{E}(L) = 2\frac{L_{J-1}^{1/2}(-1)}{L_J(-1)} = 2\frac{L_J^{1/2}(-1)}{L_J(-1)} - 2.$$

Further, using the differential equation $xL_J''(x) + (1/2 - x)L_J'(x) + JL_J(x) = 0$, we also have that

$$\begin{aligned}\frac{d}{dx}\left(x\frac{d}{dx}L_J(-x^2)\right) &= -4xL_J'(-x^2) + 4x^3L_J''(-x^2)\\ &= (-2x + 4x^3)L_J'(-x^2) + 4xJL_J(-x^2).\end{aligned}$$

This yields

$$\mathrm{Var}(L) = 4J - \mathbb{E}(L) - \mathbb{E}(L)^2,$$

and so asymptotics of the variance follow from those for the mean.

Next introduce a version of Perron's formula (see [5]),

$$L_n^\alpha(-1) = \frac{1}{2\sqrt{\pi e}}m^{\alpha/2-1/4}e^{2\sqrt{m}}\left(1 + C_1(\alpha)m^{-1/2} + C_2(\alpha)m^{-1} + \mathrm{O}(m^{-3/2})\right),$$

where $m = n + 1$ and $C_j(\alpha)$ are known explicitly. In particular, $C_1(1/2) = -1/6$, $C_2(1/2) = -7/144$, $C_1(-1/2) = -2/3$, and $C_2(-1/2) = 77/144$. Substituting into the above we then obtain

$$\mathbb{E}(L) = 2\sqrt{J+1} - 1 - \frac{2}{3\sqrt{J+1}} + \mathrm{O}(J^{-1}) = 2\sqrt{J} - 1 + \frac{1}{3\sqrt{J}} + \mathrm{O}(J^{-1}),$$

and $\mathrm{Var}(L) = 2\sqrt{J} - \frac{4}{3} + O(J^{-1/2})$ which completes the proof of point 1 (recall $J = N/2$).

Moving to the limit law for L, we introduce the notation

$$p_N(k) = \frac{C_N}{\Gamma(\frac{N}{2} - \frac{k}{2} + 1)\Gamma(\frac{k}{2} + \frac{1}{2})\Gamma(\frac{k}{2} + 1)} = C_N q_N(k)^{-1}$$

with $C_N = \Gamma(\frac{N}{2} + \frac{1}{2})[L_{N/2}(-1)]^{-1}$. For k even, $p_N(k)$ is the probability of k particles of charge 1, otherwise this probability is zero; this is just a rewrite of point 1 of Theorem 2.1 using the Gamma function. In the continuum limit this distinction is unimportant; we will show that, as $N \to \infty$

$$(2N)^{1/4}\, p_N\Big(\Big\lfloor (2N)^{1/2} + (2N)^{1/4} c \Big\rfloor\Big) = \frac{e^{-c^2/2}}{\sqrt{2\pi}}(1 + O(N^{-1/4})) \tag{4.1}$$

uniformly for c on compact sets.

First note that by Stirling's approximation (in the form $\Gamma(z) = \sqrt{\frac{2\pi}{z}}(z/e)^z(1 + O(\frac{1}{z})))$ and again Perron's formula (now in the simpler form $L_z(-1) = \frac{1}{2\sqrt{\pi e z}} e^{2\sqrt{z}}(1 + O(\frac{1}{\sqrt{z}})))$,

$$C_N = 2\pi\sqrt{Ne}\,(N/2)^{(N/2)} e^{-N/2-\sqrt{2N}}(1 + O(N^{-1/2})). \tag{4.2}$$

Next, with both k and $N - k$ large we have

$$q_N(k) = (2\pi)^{3/2}\sqrt{Nk}\,(N/2)^{(N/2)} e^{-N/2-k/2} \times e^{[(N/2-k/2)\log(1-k/N)+(k/2)\log(k^2/2N)]}(1+O(k^{-1} \vee (N-k)^{-1} \vee kN^{-1})), \tag{4.3}$$

again by Stirling's approximation. Restricting to $k = O(\sqrt{N})$, (4.2) and (4.3) yield

$$p_N(k) = \sqrt{\frac{e}{2\pi k}} e^{-\phi_N(k)}(1 + O(N^{-1/2})), \tag{4.4}$$

where

$$\phi_N(k) = \sqrt{2N} - \frac{k}{2} + \left(\frac{N}{2} - \frac{k}{2}\right)\log\left(1 - \frac{k}{N}\right) + \frac{k}{2}\log\left(\frac{k^2}{2N}\right).$$

Now, quite simply

$$\left(\frac{N}{2} - \frac{k}{2}\right)\log\left(1 - \frac{k}{N}\right) = -\frac{k}{2} + \frac{k^2}{4N} + O(N^{-1/2}),$$

if $k = O(\sqrt{N})$, and, if k is also such that $1 - \frac{k^2}{2N} = O(N^{-1/4})$, we further have

$$\frac{k}{2}\log\left(\frac{k^2}{2N}\right) = -\frac{k}{2}\left(1 - \frac{k^2}{2N}\right) - \frac{k}{4}\left(1 - \frac{k^2}{2N}\right)^2 + O(N^{-1/4}).$$

More precisely, from the last two displays we readily find that

$$\phi_N(\sqrt{2N} + \ell) = \frac{1}{2} + \frac{\ell^2}{2\sqrt{2N}} + O(N^{-1/4}), \quad \text{uniformly for } \ell = O(N^{1/4}).$$

Substituting back into (4.4), since $(\sqrt{2N}+\ell)^{-1/2} = (2N)^{-1/4}(1+O(N^{-1/4}))$ again for $\ell = O(N^{-1/4})$, completes the verification of (4.1).

Last, for the tail estimate, revisiting (4.2) and (4.3) shows the conclusion of (4.4) may be modified to read

$$C^{-1}k^{-1/2}e^{-\phi_N(k)} \leq p_N(k) \leq Ce^{-\phi_N(k)},$$

for all $1 \leq k \leq N$ with a numerical constant C. (Here we understand $(1-\frac{k}{N})\log(1-\frac{k}{N})$ to be zero at $k = N$.) Differentiating yields

$$\frac{d}{dk}\phi_N(k) = \frac{1}{2}\log\left(\frac{k^2}{2(N-k)}\right),$$

and so $\phi_N(k)$ is decreasing for $k < c^{-1}\sqrt{2n}$ and increasing for $k > c\sqrt{2N}$ for any $c > 1$. Now, since $(1-\epsilon)\log(1-\epsilon) \geq -\epsilon$ and $\log(1+\epsilon) \geq \epsilon - \epsilon^2/2$ for $0 < \epsilon \leq 1$,

$$\phi_N((1+\epsilon)\sqrt{2N}) \geq -\epsilon\sqrt{2N} + 2(1+\epsilon)\sqrt{2N}\log(1+\epsilon) \geq \epsilon\sqrt{2N},$$

also for $0 < \epsilon \leq 1$. Hence, for $c > 1$, $\mathrm{Prob}(L > c\sqrt{2N}) \leq Np_N(c\sqrt{2N}) \leq CNe^{-((c-1)\wedge 1)\sqrt{2N}}$. The proof for the left tail is much the same.

4.3 Proof of Theorem 2.3

In both cases we use the expression of the one point function in terms of Hermite polynomials, see (4.33) and (4.36) below.

We start with

$$s_1^{(N)}(x) = \sqrt{2}\sum_{n=0}^{N/2-1}\frac{\epsilon_1\widetilde{p}_{2n+1}(\sqrt{N}x)\widetilde{p}_{2n}(\sqrt{N}x) - \widetilde{p}_{2n+1}(\sqrt{N}x)\epsilon_1\widetilde{p}_{2n}(\sqrt{N}x)}{r_n},$$

and

$$s_2^{(N)}(x) = \frac{2}{\sqrt{N}}\sum_{n=0}^{N/2-1}\frac{\widetilde{p}'_{2n+1}(\sqrt{N}x)\widetilde{p}_{2n}(\sqrt{N}x) - \widetilde{p}_{2n+1}(\sqrt{N}x)\widetilde{p}'_{2n}(\sqrt{N}x)}{r_n},$$

along with the relations $\int_{-\infty}^{x} \widetilde{p}_{2n+1} = \epsilon_1 \widetilde{p}_{2n+1}(x) = 2\widetilde{p}_{2n}(x)$ and $\widetilde{p}_{2n}'(x) = \epsilon_4 \widetilde{p}_{2n}(x) = -\frac{1}{2}\widetilde{p}_{2n+1}(x)$. An integration by parts in both instances then allows: with $t_N = t/\sqrt{N}$,

$$
\begin{aligned}
&\int_{-\infty}^{\infty} e^{itx} s_1^{(N)}(x)dx \\
&\quad = \frac{4\sqrt{2}}{\sqrt{N}} \sum_{n=0}^{N/2-1} r_n^{-1} \int_{-\infty}^{\infty} e^{it_N x} (\widetilde{p}_{2n}(x))^2 dx \\
&\qquad - \frac{2\sqrt{2} i t_N}{\sqrt{N}} \sum_{n=0}^{N/2-1} r_n^{-1} \int_{-\infty}^{\infty} e^{it_N x} \widetilde{p}_{2n}(x) \epsilon_1 \widetilde{p}_{2n}(x) dx,
\end{aligned} \tag{4.5}
$$

and

$$
\begin{aligned}
&\int_{-\infty}^{\infty} e^{itx} s_2^{(N)}(x)dx \\
&\quad = \frac{2}{N} \sum_{n=0}^{N/2-1} r_n^{-1} \int_{-\infty}^{\infty} e^{it_N x} (\widetilde{p}_{2n+1}(x))^2 dx \\
&\qquad - \frac{2 i t_N}{N} \sum_{n=0}^{N/2-1} r_n^{-1} \int_{-\infty}^{\infty} e^{it_N x} \widetilde{p}_{2n}(x) \widetilde{p}_{2n+1}(x) dx.
\end{aligned} \tag{4.6}
$$

The first, and primary, step is to show that the advertised limits stem from the first sums on the right of the above expressions.

Lemma 4.1 *Let* $\hat{s}_N^{(1)}(t)$ *and* $\hat{s}_N^{(2)}(t)$ *denote, respectively, the first term on the right hand side of* (4.5) *and* (4.6). *Then,*

$$
\hat{s}_N^{(1)}(t) \to \frac{\sin\sqrt{2}t}{\sqrt{2}t}, \quad \hat{s}_N^{(2)}(t) \to \frac{\sqrt{2}}{t} J_1(\sqrt{2}t)
$$

as $N \to \infty$.

Proof Since $\widetilde{H}_k(x) = H_k(x)e^{-x^2/2}$ are the eigenfunctions of the Fourier transform – in particular $(\widetilde{H}_k)\widehat{\ }(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{ixu} \widetilde{H}_k(u) du = i^k H_k(x)$– we have that

$$
\begin{aligned}
\widehat{\widetilde{p}_{2n}}(x) &= \sum_{k=0}^{n} (-1)^k a_k \widetilde{H}_{2k}(x), \\
\widehat{\widetilde{p}_{2n+1}}(x) &= i \sum_{k=0}^{n} (-1)^k a_k (\widetilde{H}_{2k+1}(x) + 4k \widetilde{H}_{2k-1}(x)) = 2ix \sum_{k=0}^{n} (-1)^k a_k \widetilde{H}_{2k}(x).
\end{aligned} \tag{4.7}
$$

The last equality makes use of the three term recurrence $H_{n+1}(x) = 2xH_n(x) - 2nH_{n-1}(x)$. Plancheral's identity then yields,

$$\hat{s}_N^{(1)}(t) = \frac{4\sqrt{2}}{\sqrt{N}} \sum_{n=0}^{N/2-1} r_n^{-1} \sum_{0\le k,\ell\le n} (-1)^{k+\ell} a_k a_\ell \int_{-\infty}^{\infty} \widetilde{H}_{2k}(x+t_N/2)\widetilde{H}_{2\ell}(x-t_N/2)\,dx, \tag{4.8}$$

and

$$\hat{s}_N^{(2)}(t) = \frac{8}{N} \sum_{n=0}^{N/2-1} r_n^{-1} \sum_{0\le k,\ell\le n} (-1)^{k+\ell} a_k a_\ell \int_{-\infty}^{\infty} (x^2-(t_N/2)^2)\widetilde{H}_{2k}(x+t_N/2)\widetilde{H}_{2\ell}(x-t_N/2)\,dx. \tag{4.9}$$

We begin with the asymptotic considerations of (4.8) which is slightly simpler.

From the expansion $H_n(a+b) = \sum_{k=0}^n \binom{n}{k} H_k(a)(2b)^{n-k}$ we find that

$$\int_{-\infty}^{\infty} \widetilde{H}_{2k}(x+t/2)\widetilde{H}_{2\ell}(x-t/2)\,dx = e^{-t^2/4}\sqrt{\pi} \sum_{m=0}^{2(k\wedge\ell)} \binom{2k}{m}\binom{2\ell}{m} m!(-2)^m t^{2k+2\ell-2m}.$$

Given this, $\hat{s}_N^{(1)}$ is equivalent, as $N\to\infty$, to

$$\begin{aligned} \hat{s}_{N,d}^{(1)} + \hat{s}_{N,o}^{(1)} &= \frac{4\sqrt{2\pi}}{\sqrt{N}} \sum_{n=0}^{N/2-1} r_n^{-1} \sum_{k=0}^{n} a_k^2 2^{2k}(2k)! \sum_{m=0}^{2k} \frac{(2k)_m}{(m!)^2}\left(-\frac{t^2}{2N}\right)^m \\ &+ \frac{8\sqrt{2\pi}}{\sqrt{N}} \sum_{n=1}^{N/2-1} r_n^{-1} \sum_{0\le k<\ell\le n} a_k a_\ell 2^{k+\ell}(2k)! \sum_{m=0}^{2k} \frac{(2\ell)_{2\ell-2k+m}}{m!(2\ell-2k+m)!}\left(-\frac{t^2}{2N}\right)^{\ell-k+m}, \end{aligned} \tag{4.10}$$

in which we have introduced a self-evident notation for the diagonal and off-diagonal components as well as the (nontraditional) shorthand $(n)_m := \frac{n!}{(n-m)!}$.

Next, recall the definitions $a_n = \frac{n!}{(2n)!}L_n^{(-1/2)}(-1)$, $r_n = \sqrt{\pi}2^{2n+2}(2n+2)!a_n a_{n+1}$ and note the simple appraisals: with $c = (4\pi e)^{-1}$,

$$\begin{aligned} r_n^{-1} &= \frac{1}{4c\pi\sqrt{n}} e^{-4\sqrt{n}}(1+O(n^{-1/2})), \\ a_n a_{n+m} 2^{2n+2m}(2n)! &= c\sqrt{\pi} n^{-m-1/2} e^{4\sqrt{n}}(1+O(mn^{-1/2})), \end{aligned} \tag{4.11}$$

where the latter will be used for m nonnegative and moderate (compared with $n^{1/2}$). We will also make repeated use of the fact

$$\sum_{k=1}^{n} n^{m-1/2} e^{4\sqrt{n}} = \frac{1}{2} n^m e^{4\sqrt{n}}(1+O(n^{-1/2})), \tag{4.12}$$

valid for any real m.

Continuing, we change the order of summation to write

$$\hat{s}_{N,d}^{(1)} = \frac{4\sqrt{2\pi}}{\sqrt{N}} \sum_{m=0}^{N-2} \frac{(-t^2/2N)^m}{(m!)^2} \sum_{n=\lceil m/2 \rceil}^{N/2-1} r_n^{-1} \sum_{k=\lceil m/2 \rceil}^{n} (2k)_m a_k^2 2^{2k} (2k)!. \quad (4.13)$$

Then, for fixed m,

$$\sum_{n=\lceil m/2 \rceil}^{N/2-1} r_n^{-1} \sum_{k=\lceil m/2 \rceil}^{n} (2k)_m a_k^2 2^{2k} (2k)! = \frac{N^{m+1/2}}{4\sqrt{2\pi}(2m+1)}(1+o(1)),$$

by (4.11) and (4.12), and a dominated convergence argument yields

$$\lim_{N\to\infty} \hat{s}_{N,d}^{(1)} = \sum_{m=0}^{\infty} \frac{(-t^2/2)^m}{(m!)^2(2m+1)} = \int_0^1 J_0(\sqrt{2}tx)dx, \quad (4.14)$$

for the diagonal contribution. Next, for the off-diagonal terms (second line of (4.10)), we again change the order of summation and have that

$$\hat{s}_{N,o}^{(1)} = \frac{8\sqrt{2\pi}}{\sqrt{N}} \sum_{q=1}^{N/2-1} \sum_{m=0}^{N-2q-2} \frac{(-t^2/2N)^{q+m}}{m!(2q+m)!}$$
$$\times \sum_{n=q+\lceil m/2 \rceil}^{N/2-1} r_n^{-1} \sum_{k=\lceil m/2 \rceil}^{n-q} (2k+2q)_{2q+m} a_k a_{k+q} (2k)! 2^{2k+q}. \quad (4.15)$$

With now q and m fixed,

$$\sum_{n=q+\lceil m/2 \rceil}^{N/2-1} r_n^{-1} \sum_{k=\lceil m/2 \rceil}^{n-q} (2k+2q)_{2q+m} a_k a_{k+q} (2k)! 2^{2k+q}$$
$$= \frac{2^{q+m}}{4\sqrt{\pi}} \sum_{n=1}^{N/2} \frac{1}{\sqrt{n}} e^{-4\sqrt{n}} \sum_{k=1}^{n} k^{q+m-1/2} e^{4\sqrt{k}} (1+o(1))$$
$$= \frac{N^{q+m+1/2}}{4\sqrt{2\pi}(2q+2m+1)}(1+o(1)),$$

again by (4.11) and (4.12). Hence, for bounded t,

$$\hat{s}_{N,o}^{(1)} = 2 \sum_{q=1}^{N/2-1} \sum_{m=0}^{N-2q-2} \frac{(-t^2/2)^{q+m}}{m!(2q+m)!(2q+2m+1)}(1+o(1))$$
$$= 2 \sum_{\ell=1}^{N} \frac{(-t^2/2)^\ell}{(2\ell+1)} \sum_{q=1}^{\ell} \frac{1}{(\ell-q)!(\ell+q)!}(1+o(1))$$
$$= \sum_{\ell=1}^{N} \frac{(-t^2/2)^\ell}{(2\ell+1)} \left(\frac{2^{2\ell}}{(2\ell)!} - \frac{1}{(\ell!)^2} \right)(1+o(1)),$$

after changing variables and the order of summation in line two. That is, $\hat{s}^{(1)}_{N,o}$ tends to $\frac{\sin\sqrt{2}t}{\sqrt{2}t} - \int_0^1 J_0(\sqrt{2}tx)dx$, which, combined with (4.14), proves the first statement of the lemma.

Turning to (4.9), the preceding shows that, asymptotically, the $(x^2-(t_N/2)^2)$ within the integrand may be replaced by $\frac{1}{4}H_2(x) = x^2 - 1/2$ for which there is the related evaluation: assuming $k \le \ell$,

$$\int_{-\infty}^{\infty} H_2(x)H_{2k}(x+t/2)H_{2\ell}(x-t/2)e^{-x^2}\,dx$$

$$= 4\sqrt{\pi} \sum_{\substack{0\le n\le 2k \\ n-m=0,\pm 2}} \binom{2k}{n}\binom{2\ell}{m}(-1)^m \frac{n!m!2^{\frac{n+m}{2}}t^{2k+2\ell-n-m}}{(\frac{n-m}{2}+1)!(\frac{m-n}{2}+1)!(\frac{n+m}{2}-1)!}. \tag{4.16}$$

The resulting diagonal term (when $k=\ell$ in (4.9)) then reads

$$\hat{s}^{(2)}_{N,d} = \frac{8\sqrt{\pi}}{N}\sum_{n=1}^{N/2-1} r_n^{-1}\sum_{k=1}^{n} a_k^2 2^{2k}(2k)!\sum_{m=0}^{2k}\frac{(2k)_m(2k-m)}{(m!)^2}\left(-\frac{t^2}{2N}\right)^m$$
$$-\frac{8\sqrt{\pi}}{N}\sum_{n=1}^{N/2-1} r_n^{-1}\sum_{k=1}^{n} a_k^2 2^{2k}(2k)!\sum_{m=0}^{2k-2}\frac{(2k)_{m+2}}{m!(m+2)!}\left(-\frac{t^2}{2N}\right)^{m+1}. \tag{4.17}$$

This object does not converge on its own; cancellations from the off-diagonals are required.

With similar notation to the above we decompose $\hat{s}^{(2)}_{N,o}$ as in $\sum_{p\ge 1}\hat{s}^{(2)}_{N,(o,+p)}$ in which $\hat{s}^{(2)}_{N,(o,+p)}$ is arrived at by choosing $\ell = k+p$ in (4.9). Writing out the $p=1$ case in full we have that

$$\hat{s}^{(2)}_{N,(o,+1)} = -\sum_{(N,n,k)}\sum_{m=0}^{2k}\frac{(2k+2)_{m+2}}{(m!)^2}\left(-\frac{t^2}{2N}\right)^m$$
$$+2\sum_{(N,n,k)}\sum_{m=0}^{2k-1}\frac{(2k+2)_{m+2}(2k-m)}{m!(m+2)!}\left(-\frac{t^2}{2N}\right)^{m+1}$$
$$-\sum_{(N,n,k)}\sum_{m=0}^{2k-2}\frac{(2k+2)_{m+4}}{m!(m+4)!}\left(-\frac{t^2}{2N}\right)^{m+2}. \tag{4.18}$$

Here $\sum_{(N,n,k)}$ is shorthand for $\frac{16\sqrt{\pi}}{N}\sum_{n=1}^{N/2-1}\sum_{k=1}^{n-1} r_n^{-1}a_k a_{k+1}2^{2k}(2k)!$.

Consider now the first sum on the right of (4.17) for $k = n$ only:

$$\frac{8\sqrt{\pi}}{N} \sum_{n=1}^{N/2-1} r_n^{-1} a_n^2 2^{2n} (2n)! \sum_{m=0}^{2n} (2k)_m (2n-m) \frac{(-t^2/2N)^m}{(m!)^2}$$

$$= \frac{8\sqrt{\pi}}{N} \sum_{m=0}^{N-2} \frac{(-t^2/2N)^m}{(m!)^2} \sum_{n=\lceil m/2 \rceil}^{N/2-1} r_n^{-1} a_n^2 2^{2n} (2n)! (2n)^{m+1} (1 + o(1))$$

$$= 2 \sum_{m=0}^{N} \frac{(-t^2/2)^m}{(m!)^2 (m+1)} (1 + o(1)),$$

by the same type of estimates used in the analysis of $\hat{s}_N^{(1)}$. Next, using the additional fact that $1 - 4k a_{k+1} a_k^{-1} = -k^{-1/2}(1 + O(k^{-1}))$ the remainder (or $k \le n-1$ part) of the first sum in (4.17) plus the first sum in (4.18) is asymptotic to

$$-\frac{8\sqrt{\pi}}{N} \sum_{m=0}^{N-2} \frac{(-t^2/2)^m}{(m!)^2} \sum_{n=\lceil m/2 \rceil}^{N/2} r_n^{-1} \sum_{k=\lceil m/2 \rceil}^{n-1} a_k^2 2^{2k} (2k)! \times \frac{1}{\sqrt{k}} (2k)^{m+1}$$

$$= - \sum_{m=0}^{N} \frac{(-t^2/2)^m}{(m!)^2 (m+1)} (1 + o(1)).$$

The last two displays combine to produce the advertised limit $\frac{\sqrt{2}}{t} J_1(\sqrt{2}t)$.

The above ideas propagate. In particular, the remaining terms of $\hat{s}_{N,d}^{(2)} + \hat{s}_{N,(o,+1)}^{(2)}$ balance to produce a $o(1)$ contribution, and this appraisal extends to the full sum over $p > 1$ of $\hat{s}_{N,(o,+p)}^{(2)}$. We do not reproduce the details. □

Revisiting second terms in (4.5) and (4.6) shows that the proof of Theorem 2.3 can be completed by the following (rough) overestimates.

Lemma 4.2 *As $N \to \infty$,*

$$\sum_{n=0}^{N/2-1} r_n^{-1} \int_{-\infty}^{\infty} |\widetilde{p}_{2n}(x) \epsilon_1 \widetilde{p}_{2n}(x)| dx = O(N^{1/2}),$$

$$\sum_{n=0}^{N/2-1} r_n^{-1} \int_{-\infty}^{\infty} |\widetilde{p}_{2n}(x) \widetilde{p}_{2n+1}(x)| dx = O(N).$$

Proof Along with the well known evaluation $||\widetilde{H}_n||_{L^2} = \pi^{1/4} 2^{n/2} \sqrt{n!}$ used several time already, it holds that $||\widetilde{H}_n||_{L^1} = c\, 2^{n/2} \sqrt{n!}\, n^{-1/4} (1 + O(n^{-1})$ with a (known) numerical constant c, see [10]. Next note that $||f \epsilon_1 g||_{L^1} \le ||f||_{L^1} ||g||_{L^1}$ and so

$$||\widetilde{p}_{2n}\epsilon_1\widetilde{p}_{2n}||_{L^1} \le ||\widetilde{p}_{2n}||_{L^1}^2 = \Big(\sum_{k=0}^{n} a_k ||\widetilde{H}_{2k}||_{L^1}\Big)^2$$
$$\le C\Big(\sum_{k=0}^{n}\frac{1}{\sqrt{k}}e^{2\sqrt{k}}\Big)^2 = O(e^{4\sqrt{n}}).$$

Here we have used, again, that $a_k 2^k\sqrt{2k!} \sim k^{-1/4}e^{2\sqrt{k}}$. Recalling that $r_n^{-1} \sim n^{-1/2}e^{4\sqrt{n}}$ finishes the first part.

For the second estimate we can get by with an application of Schwarz's inequality. Simply compute

$$||\widetilde{p}_{2n}||_{L^2}^2 = \sum_{k=0}^{n} a_k^2 ||\widetilde{H}_{2k}||_{L^2}^2 \le C\sum_{k=0}^{n}\frac{1}{\sqrt{k}}e^{4\sqrt{k}} = O(e^{4\sqrt{n}}),$$

and

$$||\widetilde{p}_{2n+1}||_{L^2}^2 = \sum_{k=0}^{n} a_k^2(||\widetilde{H}_{2k+1}||_{L^2}^2 + 16k^2||\widetilde{H}_{2k-1}||_{L^2}^2) + 2\sum_{k=0}^{n-1} a_k a_{k-1}||\widetilde{H}_{2k+1}||_{L^2}^2$$
$$\le C\sum_{k=0}^{n}\sqrt{k}e^{4\sqrt{k}} = O(ne^{4\sqrt{n}}),$$

to find that $r_n^{-1}||\widetilde{p}_{2n}||_{L^2}||\widetilde{p}_{2n+1}||_{L^2} = O(1)$, which suffices. □

4.4 Proofs of Theorem 2.4 and 2.5

For the asymptotic mean and variance what is important is that the statement of Theorem 2.1 is easily adjusted to account for general fugacities. In particular, for the ensemble with fugacity $= X$, $\mathrm{Prob}(L, M)$ is the coefficient of c^L in $L_{N/2}(-X^2c^2)/L_{N/2}(-X^2)$. It follows that,

$$E[L] = 2X^2\left(\frac{L_{N/2}^{1/2}(-X^2)}{L_{N/2}^{-1/2}(-X^2)} - 1\right), \quad \mathrm{Var}[L] = 2X^2N + (1-2X^2)E[L] - E[L]^2.$$

Using $L_n^{-1/2}(t^2) = (-4)^{-n}(n!)^{-1}H_{2n}(t)$, $L_n^{1/2}(t^2) = (-4)^{-n}(2n!t)^{-1}H_{2n+1}(t)$ (see also (4.34) for related identities), along with the standard representation $H_k(t) = e^{t^2}\int(-is)^k e^{its-s^2/4}\frac{ds}{\sqrt{2\pi}}$, we can write for the regime of interest that

$$E[L] = 2N\gamma\left(\frac{\int_{-\infty}^{\infty} t^{N+1}e^{N\gamma(2t-t^2)}dt}{\int_{-\infty}^{\infty} t^N e^{N\gamma(2t-t^2)}dt} - 1\right). \tag{4.19}$$

It remains to consider the asymptotics of

$$I(N,p)=\int_0^\infty t^p e^{N(\log t+2\gamma t-\gamma t^2)}(1+(-1)^{N+p}e^{-2N\gamma t})dt$$

for $p=0,1$, which stem from the extremal point $t^*=\frac{1}{2}+\frac{1}{2}\sqrt{1+\frac{2}{\gamma}}$ of the exponent $\log t+2\gamma t-\gamma t^2$. In particular, the standard Laplace method gives $I(N,1)/I(N,0)=t^*+O(N^{-2})$ which is sharper than what is claimed for the mean. The result for the variance again follows from that for the mean.

As to the proof of Theorem 2.5, we only describe the computation for the limiting charge 1 density. Considerations for the charge 2 density are much the same. Furthermore, since $s_1^{(N,\gamma)}$ is structurally identical to $s_1^{(N)}$ – the only differences are the scaling parameters and the dependencies of the various weights (a_n's, r_n's) on $N\gamma$ – a formula like (4.5) still holds for $\int e^{itx}s_1^{(N,\gamma)}(x)dx$. In addition, for similar reasons to the above (recall, Lemma 4.2) the second term in this corresponding formula vanishes in the $N\to\infty$ limit. That is to say, to determine the limit of the characteristic function of $s_1^{(N,\gamma)}$ it is sufficient to take the limit of

$$\begin{aligned}\hat{s}_{N,\gamma}^{(1)}(t)&=2\sum_{m=0}^{N-2}\frac{(-t^2/2)^m}{(m!)^2}\sum_{n=\lceil m/2\rceil}^{N/2-1}\sum_{k=\lceil m/2\rceil}^{n}\frac{(2k)_m}{N^m}\nu(k,0)\\&\quad+4\sum_{q=1}^{N/2-1}\sum_{m=0}^{N-2q-2}\frac{(-t^2/2)^{q+m}}{m!(2q+m)!}\sum_{n=q+\lceil m/2\rceil}^{N/2-1}\sum_{k=\lceil m/2\rceil}^{n-q}\frac{(2k+2q)_{2q+m}}{N^{q+m}}\nu(k,q).\end{aligned}\tag{4.20}$$

Here

$$\nu(k,q)=4\gamma\sqrt{\pi}\,r_n^{-1}\,a_k a_{k+q}(2k)!2^{2k+q},$$

compare (4.13) and (4.15). Now $r_k=r_k(N\gamma)$, $a_k=a_k(N\gamma)$, and it is convenient to write ν in the form

$$\begin{aligned}\nu(k,q)=&\frac{1}{\sqrt{\pi}N}\frac{n!}{\Gamma(n+\frac{1}{2})}\left(\binom{2k+2q}{k+q}^{-1}2^{2k+2q}\right)\frac{k!2^{-q}}{(k+q)!}\\&\times\frac{(n!)^2}{k!(k+q)!}(N\gamma)^{2(k-n)+q}\frac{I_k(N\gamma)I_{k+q}(N\gamma)}{I_n(N\gamma)I_{n+1}(N\gamma)}\end{aligned}\tag{4.21}$$

in which $I_k(N\gamma)=\int t^{2k}e^{N\gamma(2t-t^2)}dt$. We have simply put all appearances of $L_k(-N\gamma)$'s appearing in the a_k's, r_k's in the same type of integral representation used for the asymptotics of the mean, see (4.19).

Next, for the asymptotics of (4.20) one makes the change of variables $k = n - \ell$, intuitively viewing n as large and ℓ moderate. In that regime, the integrals $I_{n-\ell}(N\gamma)$ concentrate around the point

$$t^*(x) = \frac{1}{2} + \frac{1}{2}\sqrt{1+4x}, \quad x = \frac{n}{N\gamma}, \tag{4.22}$$

(note the obvious connection to the extremal point t^* occurring in the mean calculation), and it holds

$$\frac{I_{n-\ell}(N\gamma)}{I_n(N\gamma)} = \left(t^*(\frac{n}{N\gamma})\right)^{-2\ell}(1+o(1)),$$

for n, N large compared with ℓ. Returning to (4.21) we then have, in the same regime (and fixed q):

$$\nu(n-\ell, q) = \frac{1}{N}(2n)^{-q}\left(\frac{n}{N\gamma}\right)^{2\ell-q}\left(t^*(\frac{n}{N\gamma})\right)^{-4\ell+2q-2}(1+o(1)). \tag{4.23}$$

Now go back to the inner sums (those over $n, k = n - \ell$ in (4.20)): for $q = 0, 1, 2, \ldots$ we have that,

$$\begin{aligned}
&\sum_{n=q+\lceil m/2\rceil}^{N/2-1}\sum_{\ell=q}^{n-\lceil m/2\rceil}\frac{(2n-2\ell+2q)_{2q+m}}{N^{q+m}}\nu(n-\ell,q)\\
&= \frac{1}{N}\sum_{n=0}^{N/2-1}\left(\frac{2n}{N}\right)^{q+m} t^*\left(\frac{n}{N\gamma}\right)^{-2}\left(\frac{(\frac{n}{N\gamma})}{t^*(\frac{n}{N\gamma})^2}\right)^{-q}\sum_{k=q}^{n}\left(\frac{(\frac{n}{N\gamma})^2}{t^*(\frac{n}{N\gamma})^4}\right)^{\ell}(1+o(1))\\
&= \frac{1}{N}\sum_{n=0}^{N/2-1}\left(\frac{2n}{N}\right)^{q+m} t^*\left(\frac{n}{N\gamma}\right)^{-2}\left(\frac{(\frac{n}{N\gamma})}{t^*(\frac{n}{N\gamma})^2}\right)^{q}\frac{1}{\left(1-\frac{(\frac{n}{N\gamma})^2}{t^*\left(\frac{n}{N\gamma}\right)^4}\right)}(1+o(1))\\
&= \int_0^{1/2}(2x)^{q+m}\left(\frac{x}{\gamma t^*(x/\gamma)}\right)^{q}\frac{dx}{t^*(x/\gamma)^2-(x/\gamma)^2/t^*(x/\gamma)^2}(1+o(1)).
\end{aligned}$$

The first line uses (4.23). The second line is just the geometric series; one may check that $x^2t^*(x)^{-4} < 1$ for $x > 0$. The final line identifies the obvious Riemann sum. To finish, a bit of algebra shows that

$$\frac{1}{t^*(x/\gamma)^2-(x/\gamma)^2/t^*(x/\gamma)^2} = \frac{1}{\sqrt{1+4x/\gamma}} = \frac{1}{u_\gamma(2x)}$$

of the statement. Likewise, $\frac{x}{\gamma t^*(x/\gamma)} = v_\gamma(2x)$, also from the statement. Then, after a self-evident change of variables, the rest of the calculation goes through just as the analogous part of the $X = 1$ case.

4.5 Proof of Theorem 3.1

We will prove something slightly more general which will be useful in the sequel.

Given measures ν_1 and ν_2 on $\mathbb{R}$, define

$$Z_{L,M}^{\nu_1,\nu_2} = \frac{1}{L!M!} \int_{\mathbb{R}^L} \int_{\mathbb{R}^M} \left\{ \prod_{j<k} |\alpha_j - \alpha_k| \prod_{m<n} |\beta_m - \beta_n|^4 \prod_{\ell=1}^{L} \prod_{m=1}^{M} |\alpha_\ell - \beta_m|^2 \right\} d\nu_1^L(\alpha)\, d\nu_2^M(\beta), \tag{4.24}$$

and

$$Z^{\nu_1,\nu_2}(X) = \sum_{(L,M)} X^L Z_{L,M}^{\nu_1,\nu_2}. \tag{4.25}$$

Theorem 4.3 *Suppose N is even and $\mathbf{p}$ is any complete family of monic polynomials.*

$$Z^{\nu_1,\nu_2}(X) = \mathrm{Pf}\left(X^2 \mathbf{A}^{\mathbf{p}} + \mathbf{B}^{\mathbf{p}}\right), \tag{4.26}$$

where

$$\mathbf{A}^{\mathbf{p}} = \left[\langle p_m | p_n \rangle_1^{\nu_1}\right]_{m,n=0}^{N-1} \qquad \text{and} \qquad \mathbf{B}^{\mathbf{p}} = \left[\langle p_m | p_n \rangle_4^{\nu_2}\right]_{m,n=0}^{N-1}.$$

4.5.1 The confluent Vandermonde determinant

A special case of the confluent Vandermonde determinant identity [12] has that

$$\det \begin{bmatrix} 1 & & 1 & 1 & 0 & & 1 & 0 \\ \alpha_1 & & \alpha_L & \beta_1 & 1 & & \beta_M & 1 \\ \alpha_1^2 & \cdots & \alpha_L^2 & \beta_1^2 & 2\beta_1 & \cdots & \beta_M^2 & 2\beta_M \\ \vdots & \ddots & \vdots & & \vdots & \ddots & \vdots & \\ \alpha_1^{N-1} & \cdots & \alpha_L^{N-1} & \beta_1^{N-1} & (N-1)\beta_1^{N-2} & \cdots & \beta_M^{N-1} & (N-1)\beta_M^{N-2} \end{bmatrix}$$
$$= \prod_{j<k} (\alpha_k - \alpha_j) \prod_{m<n} (\beta_m - \beta_n)^4 \prod_{\ell=1}^{L} \prod_{m=1}^{M} (\alpha_\ell - \beta_m)^2. \tag{4.27}$$

We will denote the matrix on the left hand side of (4.27) by $\mathbf{V}(\boldsymbol{\alpha}, \boldsymbol{\beta})$ and its determinant by $\Delta(\boldsymbol{\alpha}, \boldsymbol{\beta})$. We will later use the fact that the monomials which appear in the definition of $\mathbf{V}(\boldsymbol{\alpha}, \boldsymbol{\beta})$ can be replaced by any family of monic polynomials $\mathbf{p} = (p_0, p_2, \ldots, p_{N-1})$ with $\deg p_n = n$ without changing the determinant. We will write the resulting matrix $\mathbf{V}^{\mathbf{p}}(\boldsymbol{\alpha}, \boldsymbol{\beta})$, and we note that $\Delta(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \det \mathbf{V}^{\mathbf{p}}(\boldsymbol{\alpha}, \boldsymbol{\beta})$.

It follows from (4.24) that

$$Z_{L,M}^{\nu_1,\nu_2} = \frac{1}{L!M!} \int_{\mathbb{R}^L} \int_{\mathbb{R}^M} |\det \mathbf{V}^{\mathbf{p}}(\boldsymbol{\alpha}, \boldsymbol{\beta})| d\nu_1^L(\alpha)\, d\nu_2^M(\beta).$$

4.5.2 Notation for minors

Given a non-negative integer L, we denote the set $\{1, 2, \ldots, L\}$ by $\underline{L}$. By convention, if $L = 0$, then $\underline{L}$ is the empty set. Given a strictly increasing function $\mathfrak{t} : \underline{L} \nearrow \underline{N}$ we denote by $\mathfrak{t}'$ the unique function $\underline{N - L} \nearrow \underline{N}$ whose range is disjoint from $\mathfrak{t}$. We denote by $\mathfrak{i}$ the function $\underline{L} \nearrow \underline{N}$ which is the identity on $\underline{L}$.

We define $\operatorname{sgn} \mathfrak{t}$ as follows: Let $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_N$ be any particular basis for $\mathbb{R}^N$. We then specify that

$$\mathbf{e}_{\mathfrak{t}(1)} \wedge \cdots \wedge \mathbf{e}_{\mathfrak{t}(L)} \wedge \mathbf{e}_{\mathfrak{t}'(1)} \wedge \cdots \wedge \mathbf{e}_{\mathfrak{t}'(N-L)} = \operatorname{sgn} \mathfrak{t} \cdot \mathbf{e}_1 \wedge \mathbf{e}_2 \wedge \cdots \wedge \mathbf{e}_N.$$

That is $\operatorname{sgn} \mathfrak{t} = (-1)^k$ where k is the number of transpositions necessary to put the set

$$\mathfrak{t}(1), \cdots \mathfrak{t}(L), \mathfrak{t}'(1), \cdots, \mathfrak{t}'(N-L)$$

into order. Clearly, $\operatorname{sgn} \mathfrak{i} = 1$.

Given $\mathfrak{t} : \underline{L} \nearrow \underline{N}$ and the vector $\boldsymbol{\alpha} \in \mathbb{R}^N$, we define the vector $\boldsymbol{\alpha}_{\mathfrak{t}} \in \mathbb{R}^L$ by

$$\boldsymbol{\alpha}_{\mathfrak{t}} = \left(\alpha_{\mathfrak{t}(1)}, \alpha_{\mathfrak{t}(2)}, \ldots, \alpha_{\mathfrak{t}(L)}\right)$$

If $\mathfrak{u} : \underline{L} \nearrow \underline{N}$ is another increasing function, and $\mathbf{A} = [a_{m,n}]$ an $N \times N$ matrix, we define $\mathbf{A}_{\mathfrak{t},\mathfrak{u}}$ to be the $L \times L$ minor of $\mathbf{A}$ given by

$$\mathbf{A}_{\mathfrak{t},\mathfrak{u}} = \left[a_{\mathfrak{t}(j),\mathfrak{u}(k)}\right]_{j,k=1}^{L}.$$

4.5.3 The Laplace expansion of the determinant

Using this notation, the Laplace expansion of the determinant is given by

$$\det \mathbf{A} = \sum_{\mathfrak{t}:\underline{L}\nearrow\underline{N}} \operatorname{sgn} \mathfrak{t} \cdot \det \mathbf{A}_{\mathfrak{t},\mathfrak{i}} \cdot \det \mathbf{A}_{\mathfrak{t}',\mathfrak{i}'}.$$

In particular, the Laplace expansion of the determinant of $\mathbf{V}^{\mathbf{p}}(\boldsymbol{\alpha}, \boldsymbol{\beta})$ is given by

$$\mathbf{V}^{\mathbf{p}}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \sum_{\mathfrak{t}:\underline{L}\nearrow\underline{N}} \operatorname{sgn} \mathfrak{t} \cdot \det \mathbf{V}^{\mathbf{p}}_{\mathfrak{t},\mathfrak{i}}(\boldsymbol{\alpha}) \cdot \det \mathbf{V}^{\mathbf{p}}_{\mathfrak{t}',\mathfrak{i}'}(\boldsymbol{\beta}),$$

where the notation reflects the fact that minors of the form $\mathbf{V}^{\mathbf{p}}_{\mathfrak{t},\mathfrak{i}}$ depend only on $\boldsymbol{\alpha}$ and minors of the form $\mathbf{V}^{\mathbf{p}}_{\mathfrak{t}',\mathfrak{i}'}$ only depend on $\boldsymbol{\beta}$.

4.5.4 The total partition function

Using the previous definitions, we may write

$$|\det \mathbf{V}^{\mathbf{p}}(\boldsymbol{\alpha},\boldsymbol{\beta})| = \sum_{\mathfrak{t}:\underline{L}\nearrow\underline{N}} \operatorname{sgn}\mathfrak{t}\left\{\prod_{j<k}\operatorname{sgn}(\alpha_k-\alpha_j)\right\}\det \mathbf{V}^{\mathbf{p}}_{\mathfrak{t},\mathfrak{i}}(\boldsymbol{\alpha})\det \mathbf{V}^{\mathbf{p}}_{\mathfrak{t}',\mathfrak{i}'}(\boldsymbol{\beta}),$$

and

$$Z^{\nu_1,\nu_2}_{L,M} = \sum_{\mathfrak{t}:\underline{L}\nearrow\underline{N}} \operatorname{sgn}\mathfrak{t}\,\frac{1}{L!}\int_{\mathbb{R}^L}\left\{\prod_{j<k}\operatorname{sgn}(\alpha_k-\alpha_j)\right\}\det \mathbf{V}^{\mathbf{p}}_{\mathfrak{t},\mathfrak{i}}(\boldsymbol{\alpha})\,d\nu_1^L(\boldsymbol{\alpha})$$
$$\times\frac{1}{M!}\int_{\mathbb{R}^M}\det \mathbf{V}^{\mathbf{p}}_{\mathfrak{t}',\mathfrak{i}'}(\boldsymbol{\beta})\,d\nu_2^M(\boldsymbol{\beta}).$$

We define

$$A_{\mathfrak{t}} = \frac{1}{L!}\int_{\mathbb{R}^L}\left\{\prod_{j<k}\operatorname{sgn}(\alpha_k-\alpha_j)\right\}\det \mathbf{V}^{\mathbf{p}}_{\mathfrak{t},\mathfrak{i}}(\boldsymbol{\alpha})\,d\nu_1^L(\boldsymbol{\alpha}),$$

and

$$B_{\mathfrak{t}'} = \frac{1}{M!}\int_{\mathbb{R}^M}\det \mathbf{V}^{\mathbf{p}}_{\mathfrak{t}',\mathfrak{i}'}(\boldsymbol{\beta})\,d\nu_2^M(\boldsymbol{\beta}),$$

so that

$$Z^{\nu_1,\nu_2}_{L,M} = \sum_{\mathfrak{t}:\underline{L}\nearrow\underline{N}} \operatorname{sgn}\mathfrak{t}\cdot A_{\mathfrak{t}}B_{\mathfrak{t}'}. \tag{4.28}$$

4.5.5 Simplifying $A_{\mathfrak{t}}$

The exact integral defining $A_{\mathfrak{t}}$ appears in the calculation of the partition function of Ginibre's real ensemble. As such, we may simplify $A_{\mathfrak{t}}$ by using Lemma 4.2 of [14] to find

$$A_{\mathfrak{t}} = \operatorname{Pf}\mathbf{A}^{\mathbf{p}}_{\mathfrak{t}}.$$

4.5.6 Simplifying $B_{\mathfrak{t}'}$

We remark that the integral defining $B_{\mathfrak{t}'}$ is related, but not identical to an integral appearing in the calculation of the partition function of Ginibre's real ensemble. However, the maneuvers necessary to write $B_{\mathfrak{t}'}$ as a Pfaffian are similar in both cases.

We first write

$$\det \mathbf{V}^{\mathbf{p}}_{\mathfrak{t}',\mathfrak{i}'}(\boldsymbol{\beta}) = \sum_{\sigma\in S_{2M}} \operatorname{sgn}\sigma \prod_{m=1}^{M} p_{\mathfrak{t}'\circ\sigma(2m-1)-1}(\beta_m)\, p'_{\mathfrak{t}'\circ\sigma(2m)-1}(\beta_m),$$

so that

$$\begin{aligned} B_{\mathfrak{t}'} &= \frac{1}{M!}\sum_{\sigma\in S_{2M}} \operatorname{sgn}\sigma \int_{\mathbb{R}^M} \left\{ \prod_{m=1}^{M} p_{\mathfrak{t}'\circ\sigma(2m-1)-1}(\beta_m)\, p'_{\mathfrak{t}'\circ\sigma(2m)-1}(\beta_m)\right\} d\nu_2^M(\boldsymbol{\beta}) \\ &= \frac{1}{M!}\sum_{\sigma\in S_{2M}} \operatorname{sgn}\sigma \prod_{m=1}^{M} \int_{\mathbb{R}} p_{\mathfrak{t}'\circ\sigma(2m-1)-1}(\beta)\, p'_{\mathfrak{t}'\circ\sigma(2m)-1}(\beta)\, d\nu_2(\beta). \end{aligned}$$

Next we set

$$\Pi_{2M} = \{\sigma \in S_{2M} : \sigma(2m-1) < \sigma(2m) \text{ for } m = 1, 2, \ldots, M\},$$

then

$$\begin{aligned} B_{\mathfrak{t}'} &= \frac{1}{M!}\sum_{\sigma\in \Pi_{2M}} \operatorname{sgn}\sigma \prod_{m=1}^{M} \int_{\mathbb{R}} \Big[p_{\mathfrak{t}'\circ\sigma(2m-1)-1}(\beta)\, p'_{\mathfrak{t}'\circ\sigma(2m)}(\beta) \\ &\quad - p_{\mathfrak{t}'\circ\sigma(2m)}(\beta)\, p'_{\mathfrak{t}'\circ\sigma(2m-1)-1}(\beta)\Big]\, d\nu_2(\beta) \\ &= \operatorname{Pf} \mathbf{B}^{\mathbf{p}}_{\mathfrak{t}'}. \end{aligned}$$

4.5.7 The Pfaffian formulation for the total partition function

It follows from (4.28) that

$$Z^{\nu_1,\nu_2}_{L,M} = \sum_{\mathfrak{t}:\underline{L}\nearrow\underline{N}} \operatorname{sgn}\mathfrak{t}\cdot \operatorname{Pf}\mathbf{A}^{\mathbf{p}}_{\mathfrak{t}}\cdot \operatorname{Pf}\mathbf{B}^{\mathbf{p}}_{\mathfrak{t}'},$$

and therefore that,

$$\begin{aligned} Z^{\nu_1,\nu_2}(X) &= \sum_{(L,M)} X^L \sum_{\mathfrak{t}:\underline{L}\nearrow\underline{N}} \operatorname{sgn}\mathfrak{t}\cdot \operatorname{Pf}\mathbf{A}^{\mathbf{p}}_{\mathfrak{t}}\cdot \operatorname{Pf}\mathbf{B}^{\mathbf{p}}_{\mathfrak{t}'} \\ &= \sum_{(L,M)} \sum_{\mathfrak{t}:\underline{L}\nearrow\underline{N}} \operatorname{sgn}\mathfrak{t}\cdot X^L \operatorname{Pf}\mathbf{A}^{\mathbf{p}}_{\mathfrak{t}}\cdot \operatorname{Pf}\mathbf{B}^{\mathbf{p}}_{\mathfrak{t}'}. \end{aligned}$$

Next, we set $\mathbf{X} = X\mathbf{I}$ where $\mathbf{I}$ is the $N \times N$ identity matrix. It is clear that, for each $\mathfrak{t} : \underline{L} \nearrow \underline{N}$,

$$(\mathbf{X}\mathbf{A}^{\mathbf{p}}\mathbf{X}^{\mathsf{T}})_{\mathfrak{t}} = \mathbf{X}_{\mathfrak{t}}\mathbf{A}^{\mathbf{p}}_{\mathfrak{t}}\mathbf{X}^{\mathsf{T}}_{\mathfrak{t}},$$

and

$$\mathrm{Pf}\,(\mathbf{X}_{\mathfrak{t}}\mathbf{A}^{\mathbf{p}}_{\mathfrak{t}}\mathbf{X}^{\mathsf{T}}_{\mathfrak{t}}) = \det \mathbf{X}_{\mathfrak{t}} \cdot \mathrm{Pf}\,\mathbf{A}^{\mathbf{p}}_{\mathfrak{t}} = X^{L}\,\mathrm{Pf}\,\mathbf{A}^{\mathbf{p}}_{\mathfrak{t}}.$$

It follows that

$$Z^{\nu_1,\nu_2}(X) = \sum_{(L,M)} \sum_{\mathfrak{t}:\underline{L}\nearrow\underline{N}} \mathrm{sgn}\,\mathfrak{t} \cdot \mathrm{Pf}\,(\mathbf{X}\mathbf{A}^{\mathbf{p}}\mathbf{X}^{\mathsf{T}})_{\mathfrak{t}} \cdot \mathrm{Pf}\,\mathbf{B}^{\mathbf{p}}_{\mathfrak{t}'} = \mathrm{Pf}\,(\mathbf{X}\mathbf{A}^{\mathbf{p}}\mathbf{X}^{\mathsf{T}} + \mathbf{B}^{\mathbf{p}})$$

where the last equation comes from the formula for the Pfaffian of a sum of antisymmetric matrices [16]. Finally, since $\mathbf{X}\mathbf{A}^{\mathbf{p}}\mathbf{X}^{\mathsf{T}} = X^2\mathbf{A}^{\mathbf{p}}$, we arrive at Theorem 4.3.

4.6 Proof of Theorem 3.3

The proof of Theorem 3.3 is almost identical to the proofs of Propositions 5 and 6 in [4], we therefore will give only an outline of the salient points.

Given $x_1, \ldots, x_N, y_1, \ldots, y_N \in \mathbb{R}$ and indeterminants $a_1, \ldots, a_N, b_1, \ldots, b_N$ we define the measures η_1 and η_2 on $\mathbb{R}$ given by

$$d\eta_1(\alpha) = \sum_{n=1}^{N} a_n d\delta(\alpha - x_n) \qquad \text{and} \qquad d\eta_2(\beta) = \sum_{n=1}^{N} b_n d\delta(\beta - y_n),$$

where δ is the probability measure with unit point mass at 0. Using these measures, we will specialize the situation in Sect. 4.5 to the measures $\nu_1 = w(\mu + \eta_1)$ and $\nu_2 = w^2(\mu + \eta_2)$. We will derive a Pfaffian form for the correlation functions of the microcanonical ensemble with weight w by expanding both the integral and Pfaffian sides of (4.26) for this choice of ν_1 and ν_2 and equating coefficients of the various products of the indeterminants.

4.6.1 Expanding the integral definition of Z^{ν_1,ν_2}

Starting with (4.24), and setting $Z^{\nu_1,\nu_2} = Z^{\nu_1,\nu_2}(1)$, we have

$$\frac{Z^{\nu_1,\nu_2}(X)}{Z(X)} = \frac{1}{Z(X)} \sum_{(L,M)} \frac{X^L}{L!M!} \int_{\mathbb{R}^L}\int_{\mathbb{R}^M} \Omega_{L,M}(\boldsymbol{\alpha},\boldsymbol{\beta})\, d(\mu+\eta_1)^L(\boldsymbol{\alpha})\, d(\mu+\eta_2)^M(\boldsymbol{\beta}).$$

Expanding the products $d(\mu + \eta_1)^L(\boldsymbol{\alpha})$ and $d(\mu + \eta_2)^M(\boldsymbol{\beta})$ and relabelling the variables of integration we find

$$\frac{Z^{\nu_1,\nu_2}(X)}{Z(X)} = \frac{1}{Z(X)} \sum_{(L,M)} \sum_{\ell=0}^{L} \sum_{m=0}^{M} \frac{X^L}{\ell!(L-\ell)!m!(M-m)!} \times \int_{\mathbb{R}^L}\int_{\mathbb{R}^M} \Omega_{L,M}(\boldsymbol{\alpha},\boldsymbol{\beta})\, d\eta_1^{\ell}(\boldsymbol{\alpha}_{\mathfrak{i}})\, d\mu^{L-\ell}(\boldsymbol{\alpha}_{\mathfrak{i}'})\, d\eta_2^{m}(\boldsymbol{\beta}_{\mathfrak{i}})\, d\mu^{M-m}(\boldsymbol{\beta}_{\mathfrak{i}'}).$$

Next, we expand the products defining $d\eta_1^\ell(\boldsymbol{\alpha}_\mathfrak{i})$ and $d\eta_2^\ell(\boldsymbol{\beta}_\mathfrak{i})$ and simplify to arrive at the formula

$$\frac{Z^{\nu_1,\nu_2}(X)}{Z(X)} = \sum_{(L,M)} \sum_{\ell=0}^{L} \sum_{m=0}^{M} \sum_{\mathfrak{u}:\underline{\ell}\nearrow\underline{N}} \sum_{\mathfrak{v}:\underline{m}\nearrow\underline{N}} \left\{ \prod_{j=1}^{\ell} a_{\mathfrak{u}(j)} \prod_{k=1}^{m} b_{\mathfrak{v}(k)} \right\} R_{\ell,m}(\boldsymbol{\alpha}_\mathfrak{u}; \boldsymbol{\beta}_\mathfrak{v}). \quad (4.29)$$

That is, $Z^{\nu_1,\nu_2}/Z$ is the generating function for the correlation functions of our micro-canonical ensemble.

4.6.2 Expanding the Pfaffian formulation of Z^{ν_1,ν_2}

It is easily computed that

$$\begin{aligned}\langle f|g\rangle_1^{\nu_1} = \langle f|g\rangle_1 &+ 2\sum_{n=1}^{N} a_n \left[\widetilde{f}(x_n)\epsilon_1\widetilde{g}(x_n) - \widetilde{g}(x_n)\epsilon_1\widetilde{f}(x_n)\right] \\ &-2\sum_{n=1}^{N}\sum_{m=1}^{N} a_n a_m \widetilde{f}(x_n)\widetilde{g}(x_m)\frac{\operatorname{sgn}(x_n - x_m)}{2},\end{aligned}$$

and

$$\langle f|g\rangle_4^{\nu_2} = \langle f|g\rangle_4 + \sum_{n=1}^{N} b_n \left[\widetilde{f}(y_n)\epsilon_4\widetilde{g}(y_n) - \widetilde{g}(y_n)\epsilon_4\widetilde{f}(y_n)\right].$$

For convenience let us write

$$\mathcal{E}_{1,1}(x,y) = \frac{1}{4}\operatorname{sgn}(y-x) \qquad \text{and} \qquad \mathcal{E}_{1,2}(x,y) = \mathcal{E}_{2,1}(x,y) = \mathcal{E}_{2,2}(x,y) = 0,$$

and define

$$i(n) = \begin{cases} 1 & 1 \le n \le N; \\ 4 & N < n \le 2N, \end{cases}, \qquad X_n = \begin{cases} X & 1 \le n \le N; \\ 1 & N < n \le 2N, \end{cases},$$

$$c_n = \begin{cases} 2a_n & 1 \le n \le N; \\ b_{n-N} & N < n \le 2N, \end{cases} \qquad \text{and} \qquad z_n = \begin{cases} x_n & 1 \le n \le N; \\ y_{n-N} & N < n \le 2N. \end{cases}$$

so that,

$$\begin{aligned}X^2\langle f|g\rangle_1^{\nu_1} + \langle f|g\rangle_4^{\nu_2} = X^2\langle f|g\rangle_1 &+ \langle f|g\rangle_4 \\ &+ \sum_{n=1}^{2N} c_n X_n^2 \left[\widetilde{f}(z_n)\epsilon_{i(n)}\widetilde{g}(z_n) - \widetilde{g}(z_n)\epsilon_{i(n)}\widetilde{f}(z_n)\right] \\ &- \sum_{n=1}^{2N}\sum_{m=1}^{2N} c_n c_m X_n X_m \widetilde{f}(z_n)\widetilde{g}(z_m)\mathcal{E}_{i(m),i(n)}(z_m, z_n).\end{aligned}$$

Defining $\mathbf{A}^{\mathbf{p}}$ and $\mathbf{B}^{\mathbf{p}}$ as in Theorem 3.1 and

$$\mathbf{A}^{\mathbf{p},\nu_1} = \left[\langle p_j | p_k \rangle_1^{\nu_1}\right]_{j,k=1}^N, \qquad \mathbf{B}^{\mathbf{p},\nu_1} = \left[\langle p_j | p_k \rangle_1^{\nu_1}\right]_{j,k=1}^N,$$

we immediately see that

$$X^2 \mathbf{A}^{\mathbf{p},\nu_1} + \mathbf{B}^{\mathbf{p},\nu_2} = \underbrace{\mathbf{C}^{\mathbf{p}}}_{=X^2\mathbf{A}^{\mathbf{p}}+\mathbf{B}^{\mathbf{p}}} + \mathbf{W}^{\mathbf{p}},$$

where the j, k entry of $\mathbf{W}^{\mathbf{p}}$ is given by

$$\sum_{n=1}^{2N} c_n X_n^2 \left[\widetilde{p}_j(z_n)\epsilon_{i(n)}\widetilde{p}_k(z_n) - \widetilde{p}_k(z_n)\epsilon_{i(n)}\widetilde{p}_j(z_n)\right]$$
$$- \sum_{n=1}^{2N}\sum_{m=1}^{2N} c_n c_m X_m^2 \widetilde{p}_j(z_n)\widetilde{p}_k(z_m)\mathcal{E}_{i(m),i(n)}(z_m, z_n).$$

Next we define the $N \times 4N$ matrix $\mathbf{X}$ by

$$\mathbf{X} = \begin{bmatrix} \sqrt{c_m}\,\widetilde{p}_j(z_m) & \sqrt{c_m} X_m^2 \epsilon_{i(m)} \widetilde{p}_j(z_m) \end{bmatrix} \qquad j = 0, \ldots N-1; \quad m = 1, \ldots, 2N.$$

We also define the $4N \times 4N$ matrix $\mathbf{J}$ by

$$\mathbf{J} = \begin{bmatrix} 0 & 1 & & & \\ -1 & 0 & & & \\ & & \ddots & & \\ & & & 0 & 1 \\ & & & -1 & 0 \end{bmatrix},$$

and the $4N \times 4N$ matrix $\mathbf{Y}$ by

$$\mathbf{Y} = -\mathbf{J} + \begin{bmatrix} \sqrt{c_m c_n}\, \mathcal{E}_{i(m),i(n)}(z_m, z_n) & 0 \\ 0 & 0 \end{bmatrix}_{m,n=1}^{2N}.$$

Finally, we set $\mathbf{Z} = (\mathbf{C}^{\mathbf{p}})^{-\mathsf{T}} = \left[\zeta_{j,k}\right]_{j,k=0}^{N-1}$.

A bit of matrix algebra reveals that

$$\mathbf{A}^{\mathbf{p},\nu_1} + \mathbf{B}^{\mathbf{p},\nu_2} = \mathbf{Z}^{-\mathsf{T}} + \mathbf{X}\mathbf{Y}\mathbf{X}^{\mathsf{T}},$$

and therefore,

$$\frac{Z^{\nu_1,\nu_2}(X)}{Z(X)} = \frac{\mathrm{Pf}(\mathbf{Z}^{-\mathsf{T}} - \mathbf{X}\mathbf{Y}\mathbf{X}^{\mathsf{T}})}{\mathrm{Pf}(\mathbf{Z}^{-\mathsf{T}})}$$

This is useful, since by the Pfaffian Cauchy–Binet identity [4, 13],

$$\frac{Z^{\nu_1,\nu_2}(X)}{Z(X)} = \frac{\mathrm{Pf}(\mathbf{Y}^{-\mathsf{T}} - \mathbf{X}^{\mathsf{T}}\mathbf{Z}\mathbf{X})}{\mathrm{Pf}(\mathbf{Y}^{-\mathsf{T}})}$$

A bit more matrix algebra reveals that

$$\mathbf{X}^{\mathsf{T}}\mathbf{Z}\mathbf{X} = \begin{bmatrix} \sqrt{c_n c_m} \sum_{j,k=0}^{N-1} \widetilde{p}_j(z_m)\zeta_{j,k}\widetilde{p}_k(z_n) & \sqrt{c_n c_m} X_n^2 \sum_{j,k=0}^{N-1} \widetilde{p}_j(z_m)\zeta_{j,k}\epsilon_{i(n)}\widetilde{p}_k(z_n) \\ \sqrt{c_n c_m} X_m^2 \sum_{j,k=0}^{N-1} \epsilon_{i(m)}\widetilde{p}_j(z_m)\zeta_{j,k}\widetilde{p}_k(z_n) & \sqrt{c_n c_m} X_n^2 X_m^2 \sum_{j,k=0}^{N-1} \epsilon_{i(m)}\widetilde{p}_j(z_m)\zeta_{j,k}\epsilon_{i(n)}\widetilde{p}_k(z_n) \end{bmatrix}_{m,n=1}^{2N}.$$

And since

$$\mathbf{Y}^{-\mathsf{T}} = -\mathbf{J} - \begin{bmatrix} 0 & 0 \\ 0 & \sqrt{c_m c_n}\,\mathcal{E}_{i(m),i(n)}(z_m, z_n) \end{bmatrix}_{m,n=1}^{2N},$$

we have that $\mathrm{Pf}(\mathbf{Y}^{-\mathsf{T}}) = (-1)^N$. This implies that

$$\frac{Z^{\nu_1,\nu_2}(X)}{Z(X)} = \mathrm{Pf}\left(\mathbf{X}^{\mathsf{T}}\mathbf{Z}\mathbf{X} - \mathbf{Y}^{-\mathsf{T}}\right)$$

It follows that

$$\frac{Z^{\nu_1,\nu_2}(X)}{Z(X)} = (-1)^N\,\mathrm{Pf}(-\mathbf{J} - \mathbf{K}) = \mathrm{Pf}(\mathbf{J} + \mathbf{K}),$$

where

$$\mathbf{K} = \left[\sqrt{c_m c_n}\,K_N^{i(m),i(n)}(z_m, z_n)\right]_{m,n=1}^{2N}.$$

Using these definitions and the formula for the Pfaffian of the sum $\mathbf{J} + \mathbf{K}$, [16], we find that

$$\frac{Z^{\nu_1,\nu_2}}{Z} = \sum_{n=1}^{2N} \sum_{\mathfrak{t}:\underline{n}\nearrow\underline{2N}} \mathrm{Pf}\,\mathbf{K}_{\mathfrak{t}},$$

where $\mathbf{K}_{\mathfrak{t}}$ is the $2n \times 2n$ antisymmetric matrix given by

$$\mathbf{K}_{\mathfrak{t}} = \left[\sqrt{c_{\mathfrak{t}(j)} c_{\mathfrak{t}(k)}}\,K_N^{i\circ\mathfrak{t}(j),i\circ\mathfrak{t}(k)}(z_{\mathfrak{t}(j)}, z_{\mathfrak{t}(k)})\right]_{j,k=1}^{n}.$$

Finally, after expanding Pf $\mathbf{K}_{\mathfrak{t}}$ and unravelling the definitions of the cs and zs, we have

$$\frac{Z^{\nu_1,\nu_2}}{Z} = \sum_{\ell=0}^{N} \sum_{m=0}^{N} \sum_{\mathfrak{u}:\underline{\ell} \nearrow \underline{N}} \sum_{\mathfrak{v}:\underline{m} \nearrow \underline{N}} \left\{ \prod_{j=1}^{\ell} a_{\mathfrak{u}(j)} \prod_{k=1}^{m} b_{\mathfrak{v}(k)} \right\}$$
$$\times 2^{\ell} \operatorname{Pf} \begin{bmatrix} K_N^{1,1}(x_{\mathfrak{u}(j)}, x_{\mathfrak{u}(j')}) & K_N^{1,2}(x_{\mathfrak{u}(j)}, y_{\mathfrak{v}(k')}) \\ K_N^{2,1}(y_{\mathfrak{v}(k)}, x_{\mathfrak{u}(j')}) & K_N^{2,2}(y_{\mathfrak{v}(k)}, y_{\mathfrak{v}(k')}) \end{bmatrix}_{j,j',k,k'=1}^{\ell,\ell,m,m} \tag{4.30}$$

Comparing the coefficient of $a_1 a_2 \cdots a_\ell b_1 b_2 \cdots b_m$ in this expression with that in (4.29) we find that

$$R_{\ell,m}(\mathbf{x};\mathbf{y}) = 2^{\ell} \operatorname{Pf} \begin{bmatrix} K_N^{1,1}(x_j, x_{j'}) & K_N^{1,2}(x_j, y_k) \\ K_N^{2,1}(y_k, x_{j'}) & K_N^{2,2}(y_k, y_{k'}) \end{bmatrix}_{j,j',k,k'=1}^{\ell,\ell,m,m},$$

as desired.

4.7 Proof of Theorem 3.4

Let H_n be the standard Hermite polynomial. It is known (cf. [1])

$$\langle H_{2m}, H_{2n+1} - 4n H_{2n-1} \rangle_1 = 4 h_{2n} \delta_{m,n}, \tag{4.31}$$

where

$$h_n := \int_{\mathbb{R}} H_n^2(x) e^{-x^2} dx = \sqrt{\pi} 2^n n!.$$

Using the fact that $H_k'(x) = 2k H_{k-1}(x)$, it follows readily that

$$\begin{aligned} \langle H_{2m}, H_{2n+1} \rangle_4 &= \int_{\mathbb{R}} \left(H_{2m}(x) H_{2n+1}'(x) - H_{2m}'(x) H_{2n+1}(x) \right) dx \\ &= 2(2n+1) h_n \delta_{m,n} - 4m h_{2m-1} \delta_{m,n+1} \\ &= h_{2n+1} \delta_{m,n} - h_{2n+1} \delta_{m,n+1}. \end{aligned} \tag{4.32}$$

We look for skew orthogonal polynomials in the form of

$$P_{2m}(x) = \sum_{k=0}^{m} a_k H_{2k}(x), \qquad a_0 = 1,$$

and determine the coefficients a_k, $1 \le k \le m$, by the orthogonal conditions

$$\langle P_{2m}, H_{2k+1} - 4k H_{2k-1} \rangle = 0, \quad k = 0, 1, \ldots, m-1.$$

This gives, by (4.31) and (4.32), the following equations on a_k,

$$-h_{2k}a_{k-1} + \left[h_{2k+1} + 4(k+X^2)h_{2k}\right]a_k - h_{2k+2}a_{k+1} = 0, \quad k = 0, 1, \ldots, m-1,$$

where we define $a_{-1} = 0$. Rescale by setting $\widehat{a}_k := \frac{k!}{(2k)!}a_k$, the above equations become

$$-(k+1)\widehat{a}_{k+1} + \left(2k + \tfrac{1}{2} - X^2\right)\widehat{a}_k - \left(k + \tfrac{1}{2}\right)\widehat{a}_{k-1}, \quad k = 0, 1, \ldots, m-1,$$

which can be used to determine $\widehat{a}_k$ recursively, starting from $\widehat{a}_{-1} = 0$ and $\widehat{a}_0 = 1$. It turns out, however, that this is precisely the three-term recurrence relation for $L_k^{-\frac{1}{2}}(-X^2)$. Consequently, $\widehat{a}_k = L_k^{-\frac{1}{2}}(-X^2)$. Hence, we conclude

$$P_{2m}(x) = \sum_{k=0}^{m} a_k H_{2k}(x), \quad a_k = \frac{k!}{(2k)!}L^{-\frac{1}{2}}(-X^2). \tag{4.33}$$

The Hermite polynomials are related to the Laguerre polynomials by [17, p. 106]

$$H_{2k}(x) = (-1)^k 2^{2k} k! L_k^{-1/2}(x^2), \quad H_{2k+1}(x) = (-1)^k 2^{2k+1} k! x L_k^{-1/2}(x^2). \tag{4.34}$$

Using this relation and the facts that

$$L_k^{\alpha}(0) = \binom{k+\alpha}{k} \quad \text{and} \quad (2k)! = \frac{2^{2k}}{\sqrt{\pi}}\Gamma(k+\tfrac{1}{2})\Gamma(k+1), \tag{4.35}$$

we see that (4.33) becomes (3.6). Since P_{2m} is determined by (4.31) and (4.32), the same process shows that

$$P_{2m+1}(x) = \sum_{k=0}^{m} a_k(H_{2k+1}(x) - 4kH_{2k}(x)), \quad a_k = \frac{k!}{(2k)!}L^{-\frac{1}{2}}(-X^2). \tag{4.36}$$

Using the fact that $H_{2j}'(x) = 4jH_{2j-1}(x)$ and $H_{2j+1}(x) = 2xH_{2j}(x) - H_{2j}'(x)$, we see that

$$P_{2m+1}(x) = \sum_{k=0}^{m} a_k(2xH_{2k}(x) - 2H_{2k}'(x)) = 2xP_{2m}(x) - 2P_{2m}'(x).$$

It remains to compute $\langle P_{2m}, P_{2m+1}\rangle$. By (4.31) and (4.32), we have

$$\begin{aligned}\langle P_{2m}, P_{2m+1}\rangle &= a_m\langle P_{2m}, H_{2m+1} - 4mH_{2m-1}\rangle \\ &= a_m[-h_{2m}a_{m-1} + (h_{2m+1} + 4(m+X^2)h_{2m})a_{2m}]\end{aligned}$$

$$= a_m\sqrt{\pi}2^{m+1}m!\Big[-(2m-1)L_{m-1}^{-\frac{1}{2}}(-X^2) \\ + (4m+2X^2+1)L_m^{-\frac{1}{2}}(-X^2)\Big] \\ = a_m\sqrt{\pi}2^{m+1}m!2(m+1)L_{m+1}^{-\frac{1}{2}}(-X^2),$$

where the last step follows from the three-term relation of the Laguerre polynomials, from which (3.8) follows readily.

Finally, we turn to the proof of (3.7). Using $\frac{d}{dx}L_k^\alpha(x) = -L_{k+1}^{\alpha+1}(x)$ and $L_k^\alpha(x) = L_k^{\alpha+1}(x) - L_{k-1}^\alpha(x)$ ([17, p. 102]), (3.6) gives

$$P_{2m+1}(x) = 2x\sum_{k=0}^{m}(-1)^k\widehat{L}_k^{-\frac{1}{2}}(-X^2)\left[L_k^{\frac{1}{2}}(x^2)+L_{k-1}^{\frac{1}{2}}(x^2)\right] \\ = 2x\left[\sum_{k=0}^{m-1}(-1)^kL_k^{\frac{1}{2}}(x^2)\left(\widehat{L}_k^{-\frac{1}{2}}(-X^2)-\widehat{L}_{k+1}^{-\frac{1}{2}}(-X^2)\right)\right. \\ \left. + (-1)^m\widehat{L}_m^{-\frac{1}{2}}(-X^2)L_m^{\frac{1}{2}}(x^2)\right],$$

where $\widehat{L}_k^\alpha(x) := L_k^\alpha(x)/L_k^\alpha(0)$, the formula (3.7) then follows from

$$x\widehat{L}_k^\alpha(x) = -(\alpha+1)\left(\widehat{L}_k^\alpha(x) - \widehat{L}_{k-1}^\alpha(x)\right),$$

which is a rescaling of the identity $xL_k^{\alpha+1}(x) = -(k+1)L_{k+1}^\alpha(x) + (k+1+\alpha)L_k^\alpha(x)$ ([17, p. 102]).

Acknowledgments We are indebted to one of the anonymous referees for pointing out reference [11] and explaining to us the important physical intution which illuminates the connections, and differences, between the model(s) considered here and the real Ginibre ensemble. The authors further acknowledge the support of NSF grants DMS-0645756 (BR), DMS-0801243 (CS), and DMS-1106113 (YX).

References

1. Adler, M., Forrester, P.J., Nagao, T., van Moerbeke, P.: Classical skew orthogonal polynomials and random matrices. J. Stat. Phys. **99**(1–2), 141–170 (2000)
2. Ben Arous, G., Guionnet, A.: Large deviations for Wigner's law and Voiculescu's non-commutative entropy. Probab. Theory Relat. Fields **108**(4), 517–542 (1997)
3. Ben Arous, G., Zeitouni, O.: Large deviations from the circular law. ESAIM Probab. Stat. **2**, 123–134 (1998)
4. Borodin, A., Sinclair, C.D.: The Ginibre ensemble of real random matrices and its scaling limits. Commun. Math. Phys. **291**(1), 177–224 (2009)
5. Borwein, D., Borwein, J.M., Crandall, R.E.: Effective Laguerre asymptotics. SIAM J. Numer. Anal. **46**(6), 3285–3312 (2008)
6. Edelman, A., Kostlan, E., Shub, M.: How many eigenvalues of a random matrix are real? J. Am. Math. Soc. **7**(1), 247–267 (1994)

7. Forrester, P.J.: Log-gases and random matrices. London Mathematical Society Monographs. Princeton University Press, Princeton (2010)
8. Forrester, P.J., Nagao, T.: Eigenvalue statistics of the real Ginibre ensemble. Phys. Rev. Lett. **99** (2007)
9. Ginibre, J.: Statistical ensembles of complex, quaternion, and real matrices. J. Math. Phys. **6**, 440–449 (1965)
10. Larsson-Cohn, L.: L^p norms of Hermite polynomials and an extremal problem on Wiener chaos. Ark. Mat. **40**, 133–144 (2002)
11. Majumdar, S.N., Schehr, G.: Real roots of random polynomials and zero crossing properties of diffusion equation. J. Stat. Phys. **132**, 235–273 (2008)
12. Meray, C.: Sur un determinant dont celui de Vandermonde n'est qu'un particulier. Revue de Mathématiques Spéciales **9**, 217–219 (1899)
13. Rains, E.M.: Correlation functions for symmetrized increasing subsequences. http://arXiv.org:math/0006097 (2000)
14. Sinclair, C.D.: Averages over Ginibre's ensemble of random real matrices. Int. Math. Res. Not. **2007**, 1–15 (2007)
15. Sommers, H.-J., Wieczorek, W.: General eigenvalue correlations for the real Ginibre ensemble. J. Phys. A **41**(40) (2008)
16. Stembridge, J.R.: Nonintersecting paths, Pfaffians, and plane partitions. Adv. Math. **83**(1), 96–131 (1990)
17. Szegő, G.: Orthogonal Polynomials, vol. XXIII, 4th edn. American Mathematical Society, Providence. American Mathematical Society, Colloquium Publications (1975)